Representing Urban Forms: A Collective Learning Model with Heterogeneous Human Mobility Data

Yanjie Fu, Guannan Liu, Yong Ge, Pengyang Wang, Hengshu Zhu, Hui Xiong

Abstract—Human mobility data refers to records of human movements, such as cellphone traces, vehicle GPS trajectories, geo-tagged posts and photos. While successfully mining human mobility data can benefit many applications such as city planning, transportation, urban economics, and public safety, it is very challenging to model large-scale Heterogeneous Human Mobility Data (HHMD) that are generated from different resources. In this paper, we develop a general collective learning approach to model HHMD at an individual level towards identifying and quantifying the urban forms of residential communities. Specifically, our proposed method exploits two geographic regularities among HHMD. First, we jointly capture the correlations among residential communities, urban functions, temporal effects, and user mobility patterns by analogizing communities as documents and mobility patterns as words. Also, we further combine explicit LASSO analysis and significant testing into latent representation learning as a regularization term by analogizing compatible Point-of-Interests (POIs) as the meta-data of communities. In this way, we can learn the urban forms, including a mix of functions and corresponding portfolios, of residential communities from HHDM and POIs. We further leverage these learned results to address two application problems: real estate ranking and restaurant popularity prediction. Finally, we conduct intensive evaluations with a variety of real-world data, where experimental results demonstrate the effectiveness of our proposed modeling method and its successful applications for other problems.

Index Terms—Collective Learning, Heterogeneous Human Mobility Data, Mobility Patterns, Urban Forms

1 INTRODUCTION

With the development of Internet, mobile, and sensing technologies, big HHMD, such as cellphone traces, vehicle GPS trajectories, geo-tagged posts and photos, are increasingly available, and represent an invaluable source of information for providing decision making supports in smart growth. Successfully modeling and mining large-scale HHMD could enable us to discover generic human mobility patterns and unique urban forms in neighborhood community. The discovered knowledge can be further used for improving city planning and governance, supporting transportation engineering, improving public safety and many other purposes.

The focus of this paper is to develop a collective learning model of HHMD for the identification, quantification, and prioritization of urban forms. In this paper, we define urban form as a mixture of urban functions and corresponding portfolios in a community, as shown in Figure 1. In the pursuit of this general aim, we have three specific tasks: (i) how to combine HHMD generated from diverse sources into a uniform model space; (ii) what are the mixed urban functions of a community; (iii) how to quantify the portfolio of the mixed functions in a community. In addition, we leverage the learned results for two applications: real estate ranking and restaurant popularity prediction.

However, it is traditionally challenging to simultaneously model such HHMD in a collected way, since the data is collected from different sources to reflect distinct mobility patterns. In addition, modern communities are of mixed urban functions, and there exists a gap from human mobility directly to such mixed functions.

In this paper, we exploit two semantically-rich regularities among HHMD for quantifying urban forms. First, there are inherent correlations among residential communities, urban functions, temporal effects, and user mobility patterns. It is very important and helpful to capture the correlations while modeling the generative process of HHMD. Specifically, we introduce a latent factor, named urban form, to represent the distributions of a mix of functions in a community, and to further uncover the correlation between communities and human mobility patterns. During different time periods, we assume different functions (e.g., office,
community.

Second, POIs feature the urban functions of a community. For instance, if a lot of shopping malls, movie theaters, restaurants exist in a community, this essentially indicates the entertainment function of this community. Effectively utilizing the POI data, in addition to HHMD, could greatly help us discover the urban form of a community. While POIs in a community usually belong to a variety of categories rather than a sole one, the POI data can provide prior information about how many and which urban functions are compatible in a community. A mix of living service, entertainment, and education is indeed a compatible example, but heavy or light industry might not be compatible for community development. Therefore, we need to identify the compatible categories of POIs, and we resort to LASSO regression analysis and significant testing to find POI categories that are positively-significant to community ratings. These identified compatible categories are viewed as an alternative estimation of a mix of functions in a community. However, how to incorporate the LASSO regression analysis of POIs into the modeling of HHMD is a challenging problem.

Specifically, in this paper, we develop a general collective learning approach to model HHMD for identifying and quantifying the a mix of functions and corresponding portfolios (i.e., urban forms) of communities. Within our model, we assume there are \( K \) latent functions in urban forms and denote them with a latent categorical variable. At different time periods, a residential community exhibits different functions due to its particular portfolios of urban form. Given a specific function and a time period, a residential neighborhood has specific mobility patterns of taxi rides, bus trips, and check-ins. Here, we borrow the idea of topic modeling in text mining: we treat taxi arrival (leaving) events, bus arrival (leaving) events, and check-in events with unique spatiotemporal patterns as three types of words in three different vocabularies. Therefore, given a time period, a neighborhood has three clusters of words. We will exploit these specific human mobility patterns for identifying the latent compatible urban functions and learning the portfolios of the a mix of functions.

In this section, we first formally introduce the problem for modeling HHMD, and then provide an overview of our analytic framework.

2 Preliminaries

In this section, we first formally introduce the problem for modeling HHMD, and then provide an overview of our analytic framework.

2.1 Definitions and Problem Statement

Definition 1. (Urban Form): The urban form of a community can be defined as a set of categorical functions with a numeric weight for each function.

Definition 2. (Problem Statement): Formally, given a set of \( M \) residential communities \( E = \{ e_1, e_2, ..., e_M \} \), the goal of our problem is to model the observed HHMD in the \( M \) communities and learn the a mix of functions and corresponding portfolios so as to understand the urban forms of residential communities. For simplicity, in this study, we assume a community \( m \) has a central location (i.e., latitude and longitude) and a neighborhood area (e.g., a circle area with radius of 1 km). According to our observations, there are two geographic regularities among HHMD, which correspond to two major tasks: (1) exploiting the correlations among residential communities, urban functions, temporal effects, and human mobility patterns for modeling the generative process of HHMD, and (2) incorporate compatible POIs identified from regression analysis into the collective learning as a regularization term for enhancing the performance of learning.

2.2 Framework Overview

The collective learning framework consists of three major stages.

(1) Extracting Implicit Representation of Urban Forms via Geographical Learning. We propose to learn the urban forms (i.e., compatible urban functions and corresponding portfolios) by mining three types of mobility patterns from check-in events, taxi arriving (leaving) events, and bus arriving (leaving) events, defined next.

Definition 3. (Checkin Pattern): Given a checkin event, the checkin pattern is a triple including information about (1) checkin day, (2) checkin hour, and (3) POI category of the checkin place.

Definition 4. (Taxi Mobility Pattern): Given a taxi trajectory, we extract the leaving (i.e., pick-up) and arriving (i.e., drop-off) patterns as two tuples, each of which contains information about (1) weekday or weekend, (2) hour, and (3) leaving or arriving.
Definition 5. (Bus Mobility Pattern): Given a bus trajectory, we extract the leaving (i.e., pick-up) and arriving (i.e., drop-off) patterns as two tuples, each of which contains information about (1) weekday or weekend, (2) hour, and (3) leaving or arriving.

We then associate all these mobility patterns to a nearby residential community once their checkin, pickup or dropoff points are located within the circle area of the community with a radius of 1 km. Besides, we argue that the heterogeneous mobility patterns around a community collectively reflect the mixed functions of this area. To this end, we bring in the idea of mixture modeling and assume there are multiple latent functions in a community. Moreover, a community shows different functions during different time periods. Therefore, given a community and a time period, we can identify a unique mobility segmentation, which is defined as follows.

Definition 6. (Mobility Segment): A mobility segment is a six-item tuple including a community, a time period, a latent function of the community in this time period, checkin pattern cluster, taxi pattern cluster, and bus pattern cluster.

According to the above definition, in each mobility segment, the community has three clusters of mobility patterns generated by the urban form, including a mix of functions and corresponding portfolio, of its community. To learn the functional portfolio of each community, here we adapt the idea of topic modeling and develop a novel generative model, where the mobility patterns and clusters are analogized as words and documents, respectively, as shown in Table 1.

TABLE 1: Analog between documents and communities.

<table>
<thead>
<tr>
<th>Documents</th>
<th>Communities</th>
</tr>
</thead>
<tbody>
<tr>
<td>a word</td>
<td>the pattern of a mobility event</td>
</tr>
<tr>
<td>a document from corpus 1</td>
<td>checkin mobility events in a community</td>
</tr>
<tr>
<td>a document from corpus 2</td>
<td>taxi mobility events in a community</td>
</tr>
<tr>
<td>a document from corpus 3</td>
<td>bus mobility events in a community</td>
</tr>
</tbody>
</table>

(2) Identifying Compatible POI Categories as Explicit Representation of Urban Form. While the above model can learn compatible urban functions and corresponding portfolios, we can improve the performance and interpretability of the model by addressing two challenges: how many and which urban functions are compatible to community ratings? Since a community usually contains a variety of POIs that reflect compound urban functions, we can empirically answer the two questions by utilizing the POI data. Particularly, we first exploit consumer activities, which are encoded in mobile checkins, to rate communities. After that, we combine both LASSO analysis and significant testing to determine POI categories that are positively-significant to community ratings. The compatible POIs and corresponding frequency densities will serve as prior information and will be incorporated into the collective learning model.

(3) Combining Explicit Representation with Implicit Representation for Enhancing Collective Learning. It is more promising to combine both explicit and implicit representations for enhancing the quantification of urban forms. We propose to incorporate the Lasso regression analysis into the collective learning. Specifically, we will use compatible POIs and corresponding frequency densities (also named explicit representation) to regularize the a mix of functions and corresponding portfolios of urban form. The POI frequency densities will be placed as a prior of K a mix of functions of urban form. Based on the above idea, we will develop a regularized geographical learning model and solve the model with EM algorithms.

3 Modeling Heterogeneous Human Mobility Data

In this section, we first introduce the general idea of modeling heterogeneous human mobility data, and then introduce a collective learning model for this problem, addressing the challenges described previously.

3.1 The General Idea

There are correlations among residential communities, urban functions, temporal effects, and human mobility patterns. Therefore, in our approach, we model the generative process of checkin, taxi, and bus mobility data for each community, based on the following intuition.

Intuition 1: A mixed community is represented as a mixture of urban functions in terms of its mixed land uses, and thus forms a portfolio of a fixed set of functions.

Intuition 2: The urban functions of a mixed community change over time. For example, people may visit an area for work on workday mornings, but visit the same area for entertainment during nights and weekends.

Intuition 3: Mobility patterns reflect the functions of a community. For example, the residential function of a place can be indicated by massive leaving patterns in the early morning (e.g., people take public transit to work) and massive arriving patterns around 6PM (e.g., people go home after work). Therefore, over a certain time period, a community shows specific mobility patterns which reflect a particular urban function.
TABLE 2: The generative process of the geographic functional learning model, where Dir is a Dirichlet distribution and Cat is a multinomial distribution.

For each function \( f = k \in \{1, \ldots, K\} \):
- Draw a multinomial distribution \( \varepsilon_k \sim \text{Dir}(\varepsilon_k|\mu) \)
- Draw a multinomial distribution \( \chi_k \sim \text{Dir}(\chi_k|\vartheta) \)
- Draw a multinomial distribution \( \tau_k \sim \text{Dir}(\tau_k|\zeta) \)

For checkin latent topic \( z = q \in \{1, \ldots, Q\} \):
- Draw a multinomial distribution \( \alpha_q \sim \text{Dir}(\alpha_q|\kappa) \)

For taxi latent topic \( u = r \in \{1, \ldots, R\} \):
- Draw a multinomial distribution \( \lambda_r \sim \text{Dir}(\lambda_r|\pi) \)

For bus latent topic \( v = w \in \{1, \ldots, W\} \):
- Draw a multinomial distribution \( \delta_w \sim \text{Dir}(\delta_w|\upsilon) \)

For each community \( m \in \{1, \ldots, M\} \):
- Draw a multinomial distribution \( \eta_m \sim \text{Dir}(\eta_m|\rho) \)
- For each time period \( n \in \{1, \ldots, N\} \):
  - Draw a community function \( f \sim \text{Cat}(f|\eta_m) \)
  - For each checkin mobility pattern \( c \in \mathcal{C}_{m,n} \):
    - Draw a latent topic of checkin document \( z \sim \text{Cat}(z|\varepsilon_f) \)
  - Draw a checkin mobility pattern \( \epsilon_f \sim \text{Cat}(\epsilon_f|\alpha_z) \)
  - For each taxi mobility pattern \( t \in \mathcal{T}_{m,n} \):
    - Draw a latent topic of taxi document \( u \sim \text{Cat}(u|\chi_f) \)
    - Draw a taxi mobility pattern \( \chi_f \sim \text{Cat}(\chi_f|\lambda_r) \)
  - For each bus mobility pattern \( b \in \mathcal{B}_{m,n} \):
    - Draw a latent topic of taxi document \( v \sim \text{Cat}(v|\tau_f) \)
    - Draw a bus mobility pattern \( \tau_f \sim \text{Cat}(\tau_f|\delta_w) \)

Intuition 4: Given a time period, a community has three clusters of mobility patterns. By treating mobility patterns and clusters as words and documents, respectively, we can model the corresponding generative processes and uncover the latent urban function through topic modeling.

3.2 The General Collective Learning Framework

Figure 2 shows the graphical representation of our geographic functional learning model. Specifically, we use a multinomial distribution \( \eta_m \) over \( K \) latent functions to model the functional portfolio of the community \( m \) (Intuition 1). Based on Intuition 2, the functions of a community may vary over time. We thus segment historical mobility patterns of checkin, taxi, and bus into multiple segments in terms of \( N \) defined time periods. For example, if we define seven time periods (i.e., Monday to Sunday), we first segment mobility patterns day by day, and then group these segments into seven clusters, each of which corresponds to a day of the week. We denote a mobility segment by \( \mathcal{F} \). For each time period \( n \), a community \( m \) shows a specific urban function \( f \) drawn from \( \eta_m \). Note that each function \( f \) has: (1) a multinomial distribution \( \varepsilon_f \) over checkin latent topics, which represents the relevance of checkin latent topics to the urban function \( f \); (2) a multinomial distribution \( \chi_f \) over taxi latent topics, which represents the relevance of taxi latent topics to the urban function \( f \); and (3) a multinomial distribution \( \tau_f \) over bus latent topics, which represents the relevance of bus latent topics to the urban function \( f \) (Intuition 3). We iteratively draw: (1) a checkin latent topic \( z \) for each checkin pattern \( c \in \mathcal{C}_{m,n} \) in checkin mobility document \( \mathcal{C}_{m,n} \); (2) a taxi latent topic \( u \) for each taxi pattern \( t \in \mathcal{T}_{m,n} \) in taxi mobility document \( \mathcal{T}_{m,n} \); and (3) a bus latent topic \( v \) for each bus pattern \( b \in \mathcal{B}_{m,n} \) in bus mobility document \( \mathcal{B}_{m,n} \) (Intuition 4). In summary, Table 2 shows the generative process.

Notice that taxi, bus, and checkin mobility patterns are not independent to each other and indeed have inter-correlations. We introduce a higher-level latent variable \( f \) to simultaneously model taxi, bus, and checkin mobility patterns, which can be treated as conditional independent with each other, given the latent function \( f_m \) of the community \( m \).

3.3 An Enhanced Collective Learning Model via Lasso Regression Analysis

Even though the above framework presents a general modeling of HHMD, there is a great potential for enhancing the performance of model learning. Three additional challenges arise in achieving this goal:

- Aside from dynamic human mobility data, can we bring in the knowledge from other domains (e.g., static POIs data) as prior information for regularizing the learning of functional portfolios?
- How to determine the best number of compatible urban functions?
- How to strategically incorporate the prior knowledge into the collective learning model?

Next, we present an enhanced regularized collective learning model via Lasso significant test.

Rating Communities with Mobile Checkin Data

Location-based social networks (LBSNs) such as Foursquare and Yelp, have attracted millions of users to share their consumption events and opinions with their friends, and have enabled us to collect the checkins of local business activities from mobile Apps. To rate a community empirically, we make use of mobile check-ins data. Specifically, to estimate the density of consumer activities in the community, we first count the numbers of mobile check-in events with respect to different POI categories, denoted by \( \text{fre} \). To estimate the diversity of consumer activities in the community, we count the total number of mobile check-ins, denoted by \( \text{fre} \). While there are many categories of POIs in or around a community, we only consider the top-10 most popular categories. To this end, we combine the vibrancy, livability, and development of a community.

Identifying Compatible POI Categories and Corresponding Frequency Densities

We propose to first use Lasso regression analysis to determine the significance and the direction of the relationship between POI categories and community ratings. We consider the usual linear regression setup, for community ratings \( r \in \mathbb{R}^{M+1} \) and community POI frequency vector \( \mathbf{p} \in \mathbb{R}^{M+1} \):

\[
r = P\beta + \vartheta, \vartheta \sim \mathcal{N}(0, \sigma^2)
\]

where \( \beta \in \mathbb{R}^{M+1} \) is unknown coefficients to be estimated. While there are many categories of POIs in or around a community, two challenges regarding how many and which POI categories are positively-significant for community development need to be answered. Thus, it is critical to simultaneously select POIs while analyze the direction (i.e., positive or negative) and significance (significant or non-significant) of the selected POIs. To this end, we combine...
both Lasso regression and Chi-square testing together to tackle the above two challenges.

First, to incorporate sparsity for feature selection, we exploit the lasso estimator, defined as:

$$\beta = \arg\min_{\beta} \frac{1}{2} ||r - P\beta||^2 + \gamma ||\beta||_1$$

(2)

where $\gamma$ is a tuning parameter controlling the level of sparsity in $\beta$. After learning the Lasso regression, we use the chi-square test to examine the significance of each POI category selected by Lasso. Let $i$ be the $i$-th POI category and $P$ be a set of POI categories, the general idea is to compare the performance enhancement between the Lasso regression on $P$ and the Lasso regression on $P \cup \{i\}$, which indeed computes the ratio of the residual sum of squares (RSS) drop from Lasso regression on $P$.

$$\nu = \frac{\text{RSS}_P - \text{RSS}_{P \cup \{i\}}}{\sigma^2}$$

(3)

where $\sigma^2$ is the sample variance. This ratio of RSS drop to variance $\nu$ can be used to check the p-value and significance level of the POI category $i$ giving the cumulative chi-square distribution.

In this way, we not only select the most meaningful features, but also identify the direction (positive coefficient or negative coefficient) and the significance (p-value) of each POI category, which will be very important for setting the number of latent urban functions, and (ii) which POI categories are compatible, which will be used to compute the frequency density $v_i$ of the $i$-th compatible POI category in the community $m$ as:

$$v_i = \frac{\text{Frequency of the } i\text{-th POI category in } m}{\text{Total Number of the POIs in } m}$$

(4)

and the POI frequency density vector of $m$ is denoted as $Y_m = (v_1, \ldots, v_i, \ldots, v_c)$ where $C$ is the number of POI categories. We will regard compatible POI categories and frequency densities as an alternative estimation of compatible dimensions and corresponding portfolios of spatial structure.

### Incorporating Frequency Densities of Compatible POI Categories into Collective Learning

To incorporate the Lasso regression analysis into the collective learning, we set the number of latent urban functions (K) as the number of compatible POI categories. The compatible POI frequency density vector from Lasso analysis will be placed as a prior of the distribution of K urban functions, in order to regularize the generative process of taxi drives, bus routes, and mobile checkins.

To avoid redundant descriptions, we only introduce how to incorporate regression analysis into the collective learning. Here, $Y_m$ is the POI frequency density vector of the $m$-th community with size as $K$. Dir($\eta_m | Y_m$) is a Dirichlet distribution regularized by $Y_m$, which indeed incorporates the compatible POI frequency density distribution ($Y_m$) of the $m$-th community into the generative process of the functional portfolio of the $m$-th community ($\eta_m$). In other words, for different POI category distributions ($Y_m$), the resulting values of $\eta_m$ are distinct. Thus, the functional portfolio learned from the enhanced model is regularized and induced by both POI features and mobility patterns.

### 4 Parameter Estimation

Let us denote all parameters by $\Psi = \{\eta, \epsilon, \chi, \tau, \alpha, \lambda, \delta\}$ where $\eta = \{\eta_m\}_{m=1}^M$, $\epsilon = \{\epsilon_i\}_{i=1}^K$, $\chi = \{\chi_i\}_{i=1}^K$, $\tau = \{\tau_k\}_{k=1}^K$, $\alpha = \{\alpha_{i,k}\}_{i=1}^Q$, $\lambda = \{\lambda_{i,k}\}_{i=1}^{R \kappa}$, $\delta = \{\delta_{w,v}\}_{w,v}^W$, the hyperparameters $\Omega = \{Y, \mu, \nu, \zeta, \kappa, \varpi, \varsigma\}$, the latent assignments of functions and topics $T = \{F, Z, U, V\}$, and the observed mobility collection $D = \{C, T, B\}$ where $C = \{c_m\}_{m=1}^M$, $T = \{t_{m,n}\}_{m,n=1}^M$ and $B = \{b_{m,n}\}_{m,n=1}^M$ are the checkin, taxi, and bus mobility documents of $M$ communities for $N$ time periods, respectively. Also, we use $P_c, P_s, P_b$ to denote the vocabularies of checkin, taxi, and bus mobility patterns, respectively.

The joint distribution can be factored as

$$P(D, T, \Psi | \Omega) = P(D, T | \Psi) P(\Psi | \Omega)$$

$$= P(C | \alpha) P(\alpha | \kappa) P(T | \lambda) P(\lambda | \varpi) P(\delta | \varsigma) P(Z | \chi) P(\epsilon | \mu) P(U | \chi) P(V | \tau) P(T | \kappa) P(F | \eta) P(\eta | Y)$$

(5)

We use Collapsed Gibbs sampling for training the model. Specifically, we derive the full conditionals and posterior and obtain the update rules of both the latent assignments and the parameters. Let $C_{z,c} = \{C_{z,c} | P_{c}\}$ where $C_{z,c}$ denotes the number of checkin pattern $c$ generated by checkin latent topic $z$; $U_{u,t} = \{U_{u,t} | P_s\}$ where $U_{u,t}$ denotes the number of taxi pattern $t$ generated by latent topic $u$; $B_{b,v} = \{B_{b,v} | P_b\}$ where $B_{b,v}$ denotes the number of bus pattern $b$ generated by latent topic $v$; $Z_{f,z} = \{Z_{f,z} | P^{(m)}_v\}$ where $Z_{f,z}$ denotes the number of checkin latent topic $z$ generated by function $f$; $U_{v,t} = \{U_{v,t} | P_s\}$ where $U_{v,t}$ denotes the number of taxi latent topic $u$ generated by function $f$; $V_{f,v} = \{V_{f,v} | P^{(m)}_v\}$ where $V_{f,v}$ denotes the number of bus latent topic $v$ generated by function $f$; $P_{m,f} = \{P_{m,f} | P^{(m)}_v\}$ where $P_{m,f}$ denotes the number of mobility segments whose urban function is $f$ in a community $m$; $X^{(*)}$ represents the count of $X$ excluding the component $(*)$ (e.g., $P^{(m)}_{k,m}$ represents the count of $P_{m,k}$ excluding mobility segment $(m,n)$); $\Gamma$ denote the gamma function.

For the $n$-th mobility segment in the community $m$, the conditional posterior probability for its latent function assignment $f$ is computed by

$$P(f_{m,n} = k | D, T - f_{m,n}, \eta_m, Y) = \frac{P^{(m)}_{m,k} + Y_{mk}}{\sum_{f=1}^{K} P^{(m)}_{m,f} + Y_{mf}}$$

$$\times \prod_{z=1}^{Q} \Gamma(z_{k,z} + \mu_z) \Gamma(\sum_{z=1}^{Q} z_{k,z}^{(m)-\mu_z})$$

$$\times \prod_{v=1}^{W} \Gamma(v_{k,v} + \mu_v) \Gamma(\sum_{v=1}^{W} v_{k,v}^{(m)-\mu_v})$$

$$\times \prod_{u=1}^{R} \Gamma(u_{k,u} + \mu_u) \Gamma(\sum_{u=1}^{R} u_{k,u}^{(m)-\mu_u})$$

$$\times \prod_{k_{m,n}=1}^{W} \Gamma(k_{m,n} + \mu_{k,n}) \Gamma(\sum_{k_{m,n}=1}^{W} k_{m,n}^{(m)-\mu_{k,n}})$$

$$\times \prod_{v_{m,n}=1}^{W} \Gamma(v_{m,n} + \mu_{v,m,n}) \Gamma(\sum_{v_{m,n}=1}^{W} v_{m,n}^{(m)-\mu_{v,m,n}})$$

$$\times \prod_{k_{m,n}=1}^{W} \Gamma(k_{m,n} + \mu_{k,m,n}) \Gamma(\sum_{k_{m,n}=1}^{W} k_{m,n}^{(m)-\mu_{k,m,n}})$$

(6)

For the $i$-th checkin pattern $c_{m,n,i} \in c_{m,n}$, the conditional posterior for its latent checkin topic is computed by
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.2018.2837027, IEEE Transactions on Knowledge and Data Engineering

\[ P(z_{m,n,i} = q|D, \Psi - z_{m,n,i}) = \frac{C_{q,m,n,i} + \kappa_{q,m,n,i}}{\sum_{c=1}^{C} C_{c,m,n,i} + \kappa_c \sum_{z=1}^{Q} z_{f,m,n,z} + \mu_c}. \]  

(7)

For the \( i \)-th taxi pattern \( t_{m,n,i} \), the conditional posterior for its latent taxi topic is computed by

\[ P(u_{m,n,i} = r|D, \Psi - u_{m,n,i}) = \frac{P_{r,m,n,i} + \omega_{u,m,n,i}}{\sum_{t=1}^{T} P_{t,m,n,i} + \omega_{u,m,n,i} + \nu_r}. \]  

(8)

For the \( i \)-th bus pattern \( b_{m,n,i} \), the conditional posterior for its latent bus topic is computed by

\[ P(v_{m,n,i} = w|D, \Psi - v_{m,n,i}) = \frac{B_{w,b_m,n,i} + \zeta_{v,m,n,i}}{\sum_{b=1}^{B} B_{b,w,b_m,n,i} + \zeta_{v,m,n,i} + \zeta_v}. \]  

(9)

After all the latent assignments are learned, we obtain the update rules of the model parameters as

\[ \eta_{m,f} = \sum_{k=1}^{K} \eta_{m,k} + \eta_{m,f}, \quad \eta_{m,f} = \frac{\sum_{q=1}^{Q} z_{f,m,n,q} + \mu_q}{\sum_{z=1}^{Q} z_{f,m,n,z} + \mu_z}. \]

\[ \lambda_{u,t} = \frac{\sum_{c=1}^{C} C_{c,m,n,i} + \kappa_c \sum_{z=1}^{Q} z_{f,m,n,z} + \mu_c}{\sum_{c=1}^{C} C_{c,m,n,i} + \kappa_c \sum_{z=1}^{Q} z_{f,m,n,z} + \mu_c}. \]

5 APPLICATIONS

In this section, we show how to exploit our proposed model for two applications: (i) real estate ranking by Willingness To Pay (WTP) and (ii) examining the impact of urban forms on Willingness to Consume and Share (WCS).

5.1 Real Estate Ranking

We exploit the results of our mobility model to develop a ranker and spot communities with high willingness to pay. Empirical studies have shown that the WTP for communities can be reflected by the return rate of real estate price over a market period, i.e., rising or falling markets [5], [10], [21]. Therefore, given a market period, WTP can be measured by the ratio of the price increase relative to the starting price of a market period, i.e., \( r = \frac{P_I - P_f}{P_f} \), where \( P_f \) and \( P_I \) denote the final and initial prices, respectively. Here we follow the norms of real estate research, which typically studies rising and falling markets separately [3], [25].

To develop our ranking model, we use the portfolio of compatible urban functions extracted from the proposed model as the features of communities. We then exploit multiple linear regression to predict WTPs on a continuous scale. Formally, for the \( m \)-th community, let \( g_m \) be the benchmark WTP computed from sales prices data, \( e_m \) be the feature vector of the community, \( w \) be the parameters of multiple linear regression, \( \hat{g}_m = w^T e_m \) is the predicted return rate of the \( m \)-th community. Moreover, we combine both a point-wise ranking accuracy and a pairwise ranking accuracy as a new estimator to replace traditional least square error estimator.

\[ L_{\text{rank}} = \prod_{m=1}^{M} \frac{1}{\sigma} \exp\left(-\frac{(g_m - \hat{g}_m)^2}{2\sigma^2}\right) + \prod_{m=1}^{M} \prod_{k=1}^{M} \frac{1}{\sigma} \exp\left(-\frac{(g_m - \hat{g}_k)^2}{2\sigma^2}\right). \]  

(10)

Furthermore, we utilize a gradient descent method to find the best parameter estimate that maximizes \( L_{\text{rank}} \).

Finally, a list of communities, together with the features and WTPs, are split into two data sets, corresponding to the falling market period (from July 2011 to February 2012) and the rising market period (from February 2012 to September 2012) as shown in Figure 3, for training and testing.

5.2 Restaurant Popularity Prediction

Aside from affordable houses, we aim to exploit the learned representations of urban forms to predict restaurant popularity as the second application. In urban life, restaurants are one type of the important places for people to consume. Restaurant popularity is an indicator of community economics. By predicting the restaurant popularity, we can compare our learned urban forms with explicit feature to validate the rationale of using a latent representation learning method. In this application, we use restaurant consumption frequency as an indicator of restaurant popularity. For comparison, we extract three categories of features: explicit features (transportation and POIs that we can extract from data directly), latent features (we apply our proposed model to learn the latent features), and the combination of explicit and latent features. Then we validate the effectiveness of representation learning via prediction accuracy in machine learning models such as Elastic Net, LASSO regression, Ridge regression.

6 EVALUATION

This section details our empirical evaluation of the proposed method on real-world data.

6.1 Experiment Setup

Data Description

Table 3 shows the detailed statistics of our real-world data sets. The transportation data covers the bus system, the subway system, and the road networks of Beijing. We also extracted POI features from the Beijing POI data set. The taxi GPS traces were collected from a Beijing taxi company. Each trajectory contains trip ID, distance (m), travel time (s), average speed (km/h), pick-up time, drop-off time, pick-up point, and drop-off point. In addition, we crawled the smart card transactions from the official website of Beijing Public Transportation Group. Each bus...
trip has card ID, time, expense, balance, route name, pick-up and drop-off stop information (name, longitude, and latitude). Moreover, the Beijing check-in data were crawled from www.jiepang.com, which is a Chinese version of Foursquare. Each check-in event includes check-in time, POI name, POI category, address, longitude, latitude, and comments. Furthermore, we crawled Beijing online business reviews from www.dianping.com, which is a business review site in China. Each review contains shop ID, name, address, latitude, longitude, consumption cost, star rating (1–5), POI category, environment, service, and overall rating.

After that, we crawled the housing transactions data of all the communities in Beijing from www.soufun.com, which is the largest online real-estate system in China. Although the data sources do not cover the entire rising and falling market periods, we used the collected data to approximate the geo-mobile information of these time windows missing real-world data. This is because (1) urban infrastructures of a city change slowly in a small time period, and (2) spatiotemporal patterns of human mobility have periodicity.

### TABLE 3: Statistics of the experiment data.

<table>
<thead>
<tr>
<th>Data Sources</th>
<th>Properties</th>
<th>Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bus stop(2011)</td>
<td>Number of bus stop</td>
<td>9,470</td>
</tr>
<tr>
<td>Subway(2011)</td>
<td>Number of subway station</td>
<td>415</td>
</tr>
<tr>
<td>Road networks (2011)</td>
<td>Number of road segments</td>
<td>162,246</td>
</tr>
<tr>
<td></td>
<td>Total length(km)</td>
<td>20,202</td>
</tr>
<tr>
<td></td>
<td>Percentage of major roads</td>
<td>7.5%</td>
</tr>
<tr>
<td>POIs</td>
<td>Number of POI</td>
<td>300,511</td>
</tr>
<tr>
<td></td>
<td>Number of categories</td>
<td>13</td>
</tr>
<tr>
<td>Taxi Traces</td>
<td>Number of taxis</td>
<td>13,597</td>
</tr>
<tr>
<td></td>
<td>Effective days</td>
<td>92</td>
</tr>
<tr>
<td></td>
<td>Time period</td>
<td>APr - Aug, 2012</td>
</tr>
<tr>
<td></td>
<td>Number of trips</td>
<td>8,202,012</td>
</tr>
<tr>
<td></td>
<td>Number of GPS points</td>
<td>111,602</td>
</tr>
<tr>
<td></td>
<td>Total distance(km)</td>
<td>61,269,029</td>
</tr>
<tr>
<td>Smart Card</td>
<td>Time Period</td>
<td>Aug 2012 to May 2013</td>
</tr>
<tr>
<td>Transactions</td>
<td>Number of car holders</td>
<td>300,250</td>
</tr>
<tr>
<td></td>
<td>Number of trips</td>
<td>2,762,128</td>
</tr>
<tr>
<td>Check-Ins</td>
<td>Number of check-in POIs</td>
<td>2,762,128</td>
</tr>
<tr>
<td></td>
<td>Number of POI categories</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>Time Period</td>
<td>01/2012-12/2012</td>
</tr>
<tr>
<td>Business Review</td>
<td>Number of reviews</td>
<td>470,846</td>
</tr>
<tr>
<td></td>
<td>Number of users</td>
<td>49,280</td>
</tr>
<tr>
<td>Real Estates</td>
<td>Number of residential communities</td>
<td>2,851</td>
</tr>
<tr>
<td></td>
<td>Size of bounding box (km)</td>
<td>404</td>
</tr>
<tr>
<td></td>
<td>Time period of transactions</td>
<td>04/2011-09/2012</td>
</tr>
</tbody>
</table>

### Evaluation Metrics

To show the effectiveness of our method, we exploit the following metrics for evaluation and comparison.

**Normalized Discounted Cumulative Gain.** The discounted cumulative gain (DCG) metric is evaluated over top N communities on the ranked communities list by assuming that high-rated communities should appear earlier in the ranked list.

\[
DCG[n] = \left\{ \begin{array}{ll}
rel_1 & \text{if } n = 1 \\
DCG[n-1] + \frac{rel_n}{\log_2 n} & \text{if } n > 1
\end{array} \right.
\]

Later, given the ideal discounted cumulative gain \(DCG'\), \(NDCG\) at the \(n\)-th position can be computed as \(NDCG[n] = \frac{DCG[n]}{DCG'[n]}\), where \(rel_f\) refers to the rating of the community \(f\). The larger \(NDCG@N\) is, the higher top-\(N\) ranking accuracy is.

**Precision.** We binarize our five-level rating system (4 > 3 > 2 > 1 > 0) by treating the ratings ≥ 3 as “high-value” and ratings < 3 as “low-value”. Given a top-\(N\) community list \(E_N\) sorted in descending order of prediction values, the precision is defined as \(\text{Precision@N} = \frac{|E_N \cap E > 3|}{N}\), where \(E > 3\) are the communities whose ratings are greater or equal to 3.

### Kendall’s Tau Coefficient

Kendall’s Tau Coefficient (or Tau for short) measures the overall ranking accuracy. Let us assume that each community \(i\) is associated with a benchmark score \(y_i\) and a predicted score \(f_i\). Then, an community pair \((i, j)\) is said to be discordant, if both \(y_i > y_j\) and \(f_i > f_j\) or if both \(y_i < y_j\) and \(f_i < f_j\). Conversely, \((i, j)\) is said to be discordant, if both \(y_i < y_j\) and \(f_i > f_j\) or if both \(y_i < y_j\) and \(f_i < f_j\). Tau is given by \(\tau = \frac{h - f}{\frac{1}{2}m(m-1)}\). Diversity is defined as \(\sum_{f=1}^{K} \left( \frac{\sum_{m=1}^{M} \sum_{n=1}^{M} \log P(c_m,n|t_m,n,b_{m,n})}{\sum_{m=1}^{M} \sum_{n=1}^{M} y_{m,n} \eta_{m,f}} \right)\). The larger diversity, the better.

### Perplexity and Diversity

Perplexity and diversity are used to study parameter sensitivity, defined by \(\text{Perplexity} = \exp \left\{ -\frac{\sum_{m=1}^{M} \sum_{n=1}^{M} \log P(c_m,n|t_m,n,b_{m,n})}{\sum_{m=1}^{M} \sum_{n=1}^{M} y_{m,n} \eta_{m,f}} \right\}\), and \(\text{Diversity} = \sum_{f=1}^{K} \frac{\sum_{m=1}^{M} \sum_{n=1}^{M} y_{m,n} \eta_{m,f}}{\sum_{m=1}^{M} \sum_{n=1}^{M} y_{m,n} \eta_{m,f}}\).

### 6.2 Evaluation of Human Mobility Modeling

![Fig. 4: Performance comparison over different shapes and radius, rising market.](image)

![Fig. 5: Performance comparison over different shapes and radius, falling market.](image)

### 6.2.1 Study of Community Radius and Shapes

Generally, the radius of a community should be neither too large nor too small. If the radius is too large, a community area will overshoot another community; if the radius is too small, the coverage of a community for human mobility data is sparse. Based on the above, we pick up several candidate radius (e.g., 1 kilometer, 1.5 kilometer, 2 kilometer), as well as cycle and square shapes. Figure 4 and Figure 5 show
the performance comparison over different shapes and radius in the rising market and the falling market in the application of real estate ranking. Overall, using a circle shape and 1 kilometer as radius achieve a stable and high performance in both rising market and falling market.

**TABLE 4: Examples of temporal topics and their patterns of check-in mobility.**

<table>
<thead>
<tr>
<th>Weekday Topics</th>
<th>Weekend Topics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topic 7</td>
<td>Topic 6</td>
</tr>
<tr>
<td>R8PM</td>
<td>E10PM</td>
</tr>
<tr>
<td>R9PM</td>
<td>E10PM</td>
</tr>
<tr>
<td>R10PM</td>
<td>E9PM</td>
</tr>
<tr>
<td>R7PM</td>
<td>E10PM</td>
</tr>
<tr>
<td>R12</td>
<td>E10PM</td>
</tr>
<tr>
<td>R1PM</td>
<td>E08PM</td>
</tr>
</tbody>
</table>

Note: R, E, and S denote restaurant, entertainment, and shopping.

**TABLE 5: Examples of temporal topics and their patterns of taxi mobility.**

<table>
<thead>
<tr>
<th>Weekday Topics</th>
<th>Weekend Topics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topic 6</td>
<td>Topic 7</td>
</tr>
<tr>
<td>L08PM</td>
<td>A08AM</td>
</tr>
<tr>
<td>A08AM</td>
<td>A08AM</td>
</tr>
<tr>
<td>A08PM</td>
<td>L08AM</td>
</tr>
<tr>
<td>A06PM</td>
<td>L05PM</td>
</tr>
</tbody>
</table>

Note: L and A denote leaving and arriving patterns respectively.

**TABLE 6: Examples of temporal topics and their patterns of bus mobility.**

<table>
<thead>
<tr>
<th>Weekday Topics</th>
<th>Weekend Topics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topic 7</td>
<td>Topic 6</td>
</tr>
<tr>
<td>L08PM</td>
<td>A05PM</td>
</tr>
<tr>
<td>A08AM</td>
<td>L06PM</td>
</tr>
<tr>
<td>A05PM</td>
<td>L07AM</td>
</tr>
<tr>
<td>A05PM</td>
<td>A08AM</td>
</tr>
</tbody>
</table>

Note: L and A denote leaving and arriving patterns respectively.

6.2.2 Study of Temporal Popularity of Checkin, Taxi, and Bus Latent Topics

We compute the topic distributions of checkin, taxi, and bus with respect to different week days. Figure 9 presents the topic distributions over seven days, with values represented by color darkness. We also list the representative words for these popular topics in Tables 4, 5, and 6, respectively. Figure 9 validates that the topic distribution of mobility has a temporal pattern. First, Figure 9(a) shows that checkin latent topics 1, 3, and 4 are popular during both weekdays and weekends. This is because topics 5, 6, 7 respectively represent shopping, entertainment, and catering activities at noon or at night, as shown in Table 4. Next, Figure 9(b) shows that taxi latent topics 3 and 4 are popular only during weekdays, while topics 4 and 6 are popular during both weekdays and weekends. From Table 5, we can see topics 3 and 4 generally include arriving patterns in the morning (i.e., go to work) and leaving patterns at night (i.e., leave after work), and thus mainly happen in weekdays. Topics 6 and 7 are combinations of both working activities (i.e., arriving early in the morning and leaving after 5PM) and catering, entertainment, and commercial activities (i.e., arriving after 5PM and leaving at night), and thus are popular during both weekdays and weekends. In addition, Table 6 shows that bus latent topics 6 and 7 include both working activities as well as catering, entertainment, and commercial activities, and thus cover both weekdays and weekends. On the other hand, bus latent topic 5 with only working activities is popular on weekdays. Bus latent topic 4 is mostly about recreation activities at night and is thus popular on weekends. The above analysis demonstrates that the geographic functional learning model can capture temporal patterns of checkin, taxi, and bus mobility.

6.2.3 Study of Positively Significant POI Categories

We exploit the Lasso significant test to select the POI categories that are not just significant but also compatible to urban vibrancy, which is represented by WPL. Table 7 lists the positive and significant POI categories learned from Lasso regression analysis. We observe that it is critical to develop a vibrant community by planning a balanced mixture of catering, shopping, living, scenic spots, and education functions in a community.

**TABLE 7: Positively Significant POI Categories**

<table>
<thead>
<tr>
<th>ID</th>
<th>POI Category #</th>
<th>POI Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>Catering (e.g., restaurants, bars)</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>Shopping</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>Living</td>
</tr>
<tr>
<td>4</td>
<td>11</td>
<td>Scenic spots</td>
</tr>
<tr>
<td>5</td>
<td>14</td>
<td>Science and education</td>
</tr>
</tbody>
</table>

6.2.4 Study of Functional Distributions for High-ranked and Low-ranked Communities

Here, we visualize the functional distribution of high-ranked and low-ranked communities, and study the correlation between community rating and functional diversity. Figure 6 compares the functional distributions of high-ranked (i.e., top 1–25) and low-ranked (i.e., top 250–2530) communities. High ranked communities generally show diverse and balanced distributions among different functions, whereas low ranked communities show unbalanced distributions with low heterogeneity. This observation validates the assumption that a good functional portfolio can increase investment value.

6.2.5 Study of Rating Communities with Checkin Data

To demonstrate the rationale of using checkin data to rate communities, we examine the correlations among checkin based scores, taxi based scores, and bus based scores. Specifically, we calculated and extracted the harmonic mean scores for all communities based on taxi, bus, and checkin data.
and obtained a score vector for each data source. The correlations among taxi, bus, and check-in are estimated based on these score vectors. Figure 8 shows the correlations between three data sources: check-in data, taxi trajectories, and bus data. We can observe that check-in data has a higher positive correlation with both taxi trajectories and bus data. But, the correlations between taxi trajectories and bus data are very weak. Notice correlations represent information similarity and redundancy; the observation indicates check-in data include the information of taxi and bus, yet taxi and bus data doesn’t include the information of check-in data. Therefore, we choose the harmonic mean of check-in data for rating communities.

6.2.6 Study of Parameter Sensitivity

Here, we investigate the sensitivity of different parameter settings in terms of three metrics: likelihood, perplexity, and diversity. Figure 7(a) plots the likelihood against the number of iterations. The likelihoods in all settings converge after 100 iterations. To ensure convergence, we retrieve all the results after 200 iterations. Figure 7(b) shows that the perplexity decreases as the number of iterations decreases, in terms of different prior (\( \rho \)) settings. Since the trends of perplexity for different numbers of latent topics are similar, we only show the plots where \( Q = R = W = 10 \). Meanwhile, we notice that a smaller \( \rho \) results in a larger perplexity when \( K \) is small, and the perplexity gaps between different settings become small with the increase of \( K \). Hence, we make a trade-off and set \( \rho = 7 \) in the following experiments. In addition, when \( K \) increases from 5 to 20, the perplexity decreases smoothly. Figure 7(c) shows that the differences among the diversities in all settings are not significant, and the number of latent topics is less related with diversity. Therefore, to avoid overfitting, we set \( K = 5, Q = R = W = 7 \), because the number of time periods for mobility segments is small (i.e., \( \lambda = 7 \), one day per segment), and the sizes of vocabularies of check-in, taxi, and bus patterns are also small.

6.3 Evaluation of Real Estate Ranking Application

We evaluate the performances of exploiting the results of the proposed modeling for ranking communities with high WTPs.

Baseline Algorithms

Since real estate ranking is related to Learning-To-Rank (LTR), we compared our method against the following algorithms.

1) Coordinate Ascent [23]: uses domination loss and coordinate descent optimization.
2) LambdaMART [1]: the boosted tree version of LambdaRank. LambdaMART combines MART and LambdaRank.
3) FenchelRank [18]: designed for solving sparse learning-to-rank with an L1 constraint.
4) ListNet [2]: a listwise ranking model with permutation top-\( k \) ranking likelihood as the objective function.

Beyond traditional ranking models, we further compare with two methods specifically designed for ranking residential communities.

5) SEK [6]: exploits regression modeling, pairwise ranking objective, and sparsity regularization, to solve the residential community ranking problem. Also, its feature design includes the entropy of POI distribution, which is an summary index of functional diversity.
6) ClusRanking [8]: solves the residential community ranking problem by capturing individual, peer, and zone dependencies.
7) DivRank [7]: incorporates the diversity of the learned urban functions into the ranking objective function.

In our experiments, we used RTree to index geographic items (e.g., POIs, trajectories, checkins, etc.) and extracted the features of POIs, public transportations, and road networks. For traditional LTR algorithms, we used RankLib\(^1\). For FenchelRank, we use the source code\(^2\) provided by the author. For SEK, we set \( C = 0.01, b = 0.01 \), and \( \sigma^2 = 1000 \). For ClusRanking, we set \( \beta_1=0.8, \beta_2=25m \), latent business areas \( K = 10, \eta = 1, \mu_q = \mu_w = 0, \sigma_q = \sigma_w = 35 \) and \( M = 3 \) for hyperparameters. We randomly divided the data into 70% for training and 30% for testing, and used Matlab for result visualization.

\(^1\)http://sourceforge.net/p/lemur/wiki/RankLib/
\(^2\)http://ss.sysu.edu.cn/~py/fenchelcode.rar
Performance Comparison

We report the ranking performances of our method, compared to baseline algorithms, on the rising and falling markets, in terms of NDCG, Precision, and Tau.

Rising Market. Figure 10 shows our method performs better than the baselines over top-$k$ ranking in rising market. For example, our method offers 21%, 32.4%, 47.2% improvement in terms of NDCG@3 compared to SEK, FenchelRank, and RankBoost, respectively. Figure 10(b) shows that the top-$K$ results ($K = 3, 5, 7, 10$) of our method consist almost exclusively of estates with rating $\geq 3$. For example, all our top-10 results are high-value, compared to just 2 for random or CoordAsc ranking, and 7–8 for the best competitor.

Falling Market. As can be seen in Figure 11, our method outperforms the baselines over top-$K$ ranking by a significant margin in falling market. Specifically, our method achieves 27.5%, 17.3%, and 99% improvement in terms of NDCG@3 compared to SEK, RankBoost, and FenchelRank, respectively. We observe the overall accuracy (Tau) of our method in [7] decreases and is lower than ClusRanking and SEK. A possible explanation could be because some sources of mobility data are not collected in Falling market period. However, the Tau values of our enhanced method is high and stable in both rising and falling datasets. This shows the benefits of the automated identification of (i) the number of latent urban functions and (ii) the informative prior of POI based regularization. Finally, although our goal is to identify top investment opportunities, for completeness we also evaluate the total ranking of all estates, showing Tau scores in Table 8.

6.4 Evaluation of Restaurant Popularity Application

The goal of this evaluation is to prove that latent features learned from our proposed model can represent the functional portfolios of communities more accurately. We compare the performances of:

- **Latent Representation.** We regard the learned functional portfolios by the method proposed in our conference paper [7] (without incorporating Lasso significance testing into the representation learning) and this paper (incorporating Lasso significance testing into the representation learning) as two kinds of latent representations, denoted as Latent Features/Non-Lasso and Latent Features/Lasso respectively.

- **Explicit Features.** We extract descriptive statistics about transportation information including: taxi average velocities, taxi average commute distance, bus leaving volume, bus arriving volume, and bus stop densities, as well as POI frequencies over different categories.

- **Explicit & Latent Features.** We combine both Latent Features/Non-Lasso and explicit features, denoted as Explicit & Latent Features/Non-Lasso; and combine both Latent Features/Lasso and explicit features, denoted as Explicit & Latent Features/Lasso.

We exploit five models to predict restaurant popularity, including: Linear Regression, Ridge Regression, Lasso, Elastic Net, and Bayesian Ridge Regression. We compare the mean square error (MAE) of explicit features, latent representations (Latent Features/Non-Lasso and Latent Features/Lasso), and the combination of explicit & latent features (Explicit & Latent Features/Non-Lasso and Explicit & Latent Features/Lasso) over these five models.

Figure 12 shows the MAEs of latent representations/Lasso and the explicit & latent feature combination/Lasso are much smaller than explicit features across all these five regression models. Moreover, the performance of Explicit & Latent Features/Lasso is better than Explicit & Latent Features/Non-Lasso. These indicate that our method is improved compared with the method proposed in our conference paper [7] that doesn’t incorporate Lasso significance testing into the representation learning. Specifically,
the latent representation and the explicit & latent feature combination offer over 60% error reduction for Linear Regression and over 50% error reduction for Elastic Net. The significant improvement on MAE by the participant of latent features learned from our proposed geographical learning model, validates the fact that the geographical learning model can learn the functional portfolios more accurately and efficiently. A possible explanation is that because of the complexity of urban communities, the latent representation learned from the geographical learning model can reveal some inner natures of the structure of urban community, which cannot be depicted by explicit statistics of basic human mobilities.

6.5 Discussions

In this study, we focus on a different perspective: modeling HHMD as documents. Existing topic modeling can partially solve the problem. However, since HHMD are collected from multiple sources, traditional topical modeling (e.g., LDA) cannot simultaneously model three different types of human mobility data in the same region and at the same time period. Our proposed method addresses this challenge by jointly modeling taxi, bus, and checkin mobility data. Our model can be extended to four or more types of human mobility data.

Aside from document based modeling, HHMD can be alternatively modeled as tensors or graphs. For instance, by modeling human mobility as tensors, we can estimate the traffic volume of a specific location at a specific time period. By modeling human mobility as graphs, we can construct Intersection-to-Intersection, POI-to-POI, or Region-to-Region, to describe the pairwise correlation among locations from human movements. Unlike tensor or graph based modeling, our model is not designed for modeling the dynamics of human movements or the interactions among locations. Indeed, as a probabilistic hierarchical model, our model is very helpful for profiling, skeletonizing, and summarizing the structure information of geographic items, e.g., the urban forms including compatible urban functions and corresponding portfolios of communities in our case.

Finally, we use Gibbs sampling, which has been theoretically proved to converge, for parameter estimation. While the inference costs time, it is easy to implement and test. The efficiency can be further improved by exploiting quality computing platforms and using more vector or matrix operations in programming from an engineering perspective.

7 Related Work

There is an increasing interest in modeling and analyzing human mobility data. Below we describe some relate works that have been accomplished by researchers in different fields such as data mining, ubiquitous computing, and geographic information systems.

Researchers have developed studies for understanding human mobility data. The work published in Nature [11] indicated that human trajectories show a high degree of temporal and spatial regularity, in which each individual being characterized by a time-independent characteristic travel distance and a significant probability to return to a few highly frequented locations. Song et al. [29] explored the limits of predictability in human dynamics by studying the mobility patterns of anonymized mobile phone user and found a 93% potential predictability in user mobility across the whole user base. Meloni et al. [22] formulated and analyzed a metapopulation model that incorporates several scenarios of self-initiated behavioral changes into the mobility patterns of individuals. Yang et al. [32] found that people’s movement behaviors are strongly affected by their social interactions with each other, and presented a novel human mobility model based on heterogeneous centrality and overlapping community structure in social networks. Isaacman et al. [15] proposed a method to model how large populations move within different metropolitan areas. The method takes as input spatial and temporal probability distributions drawn from empirical mobility data, and produces synthetic mobility data for a synthetic population. The work in [4] coarsens the data spatially and temporally to devise a formula for the uniqueness of human mobility traces given their resolution, and found that the uniqueness of mobility traces decays approximately as the 1/10 power of their resolution. Giannotti et al. [9] introduced the sequential pattern mining paradigm that analyzes the trajectories of moving objects, and provided several different methods to acquire the patterns from trajectories data.

Aside from the research of understanding the nature of human mobility, there are existing studies on exploiting human mobility data for enabling various applications. The first stream is about predicting future movement [13], [31], [34]. For example, Hoang et al. [13] proposed methods to predict flows of crowds in every region of a city based on big data. Ying et al. [34] proposed a method to predict the next location of a user’s movement based on both the geographic and semantic features of users’ trajectories. The work in [31] develops a hybrid model integrating both the regularity and conformity of human mobility to make a location prediction, which captured users’ regular movement patterns and their occasional visits influenced by others. The

<table>
<thead>
<tr>
<th>Period</th>
<th>CoordAsc</th>
<th>LambdaMART</th>
<th>FenchelRank</th>
<th>SEK</th>
<th>ListNet</th>
<th>DivRank</th>
<th>Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rising</td>
<td>-0.137</td>
<td>0.072</td>
<td>0.122</td>
<td>0.349</td>
<td>0.172</td>
<td>0.343</td>
<td>0.351</td>
</tr>
<tr>
<td>Falling</td>
<td>0.223</td>
<td>0.231</td>
<td>-0.125</td>
<td>0.335</td>
<td>0.054</td>
<td>0.236</td>
<td>-0.093</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.358</td>
</tr>
</tbody>
</table>
The second stream is to exploit human mobility data for smart transportation [27], [36], [37]. Work [27] designs on detecting over-crowded stations in public transport networks by using human-tracking data, and then proposed a method based on network expansion to find unobstructed routes to go around these over-crowded stations. Zhang et al. [36] developed methods to increase both the effectiveness and the efficiency of the scheduling algorithm of the large-scale dynamic city express problem. Work [37] mines the driving route for end users by considering physical feature of a route, traffic flow, and driving behavior. The third stream is about mining human mobility data for site selection [6], [7], [8], [14], [19], [24]. The work in [19] and [14] respectively selects optimal sites for ambulance stations and air minoring stations by mining human mobility data. Niu et al. [24] mined taxi traces for gas station site selection. Work [16] selects the optimal sites for retail stores by mining Foursquare data. Work [6], [7], [8] models human mobility data for ranking highly-urban communities. The last stream is to analyze urban functions and land uses with human mobility data. Recent work studied the distributions of some geographic topics in terms of user-generated social media such as geo-tweets and geo-tagged photographing events [17], [33]. More recent studies has been conducted for identifying and quantifying urban functions and land uses by exploiting human mobility patterns [20], [35].

Finally, the proposed application in this paper is related to learning to rank with diversity. The pair-wise methods reduce the LTR task to a classification problem [12]. The goal of the pairwise ranking is to learn a binary classifier to identify the better document in a given document pair by minimizing the average number of rank inversions. Study [28] unifies both rating error and ranking error as objective function to enhance Top-K recommendation. More recent works [26], [30], [38] study diversified learning to rank. For example, [38] ranks items by random walks in an absorbing Markov chain and achieves both diversity and centrality. The work in [30] proposes a diversified ranking objective by incorporating subtopics into MAP (Mean Average Precision) for expert finding.

8 Concluding Remarks

Since human mobility patterns provide a reasonable estimation of diverse functions present in each residential community, we proposed a general collective learning model of HHMD for identification, quantification, and prioritization of urban forms (i.e., defined as compatible urban functions and corresponding portfolio of communities) via latent semantic analysis. Specifically, we provided a general modeling approach to align HHMD generated from diverse sources into a uniform model space; We develop a practical method by combining Lasso analysis and significance testing in order to answer how many and which urban functions are positively compatible for community development; We devise a latent factor that can be learned from the HHMD modeling for quantifying the portfolio of these a mix of functions in a community; We propose an analogy to strategically incorporate Lasso analysis into the modeling of HHMD. Then, we exploit the results extracted via our HHMD modeling for understanding the impact of urban forms on socioeconomic behaviors such as real estate and restaurant popularity. Finally, we conducted extensive experiments on real-world human mobility data, urban geographical data, and user check-in data collected from location based social networks. As revealed in the experimental results, a proper mixture of urban functions can be helpful for enhancing urban development and boosting local business, and the performance improvement and robustness check shows our prosed model can make good sense of human mobility data.

References


[33] Zhijun Yin, Liangliang Cao, Jiawei Han, Chengxiang Zhai, and Thomas Huang. Geographical topic discovery and comparison. In Proceedings of the 20th international conference on World wide web, pages 247–256. ACM, 2011.


Yanjie Fu received his Ph.D. degree from Rutgers University in 2016, the B.E. degree from Univ. of Sci. and Tech. of China in 2008, and the M.E. degree from Chinese Academy of Sciences in 2011. He is currently an Assistant Professor at the University of Arizona. His research interests include data mining, social networks, and business intelligence. His work has been published in the journal of Decision Support Systems, Neurocomputing, and also in the conference proceedings such as KDD, ICDM, etc.

Guannan Liu is currently an assistant professor in the department of information systems with Beihang University, Beijing, China. He received the Ph.D. degree from Tsinghua University, China. His research interests include data mining, social networks, and business intelligence. His work has been published in the journal of Decision Support Systems, Neurocomputing, etc, and also in the conference proceedings such as KDD, ICDM, etc.

Yong Ge is currently an Assistant Professor at the University of Arizona. His research interests include data mining and business analytics. He has published prolifically in refereed journals and conference proceedings, such as IEEE TKDE, ACM TOIS, ACM TKDD, ACM TIST, ACM SIGKDD. He has served as Program Committee members at the ACM SIGKDD 2013, WAIM 2013, and IEEE ICDM 2013. Also he has served as a reviewer for numerous journals, including IEEE TKDE, ACM TIST.

Pengyang Wang received the BE degree from the Xi’dian University, China, 2014, the MS degree from the Beijing University of Posts and Telecommunications, China, 2017. He is currently working toward the PhD degree in the Computer Science Department at Missouri University of Science and Technology. His research interests include data mining, business analytics, geoeconomics, and representation learning.

Hengshu Zhu is currently with Baidu Research. His general area of research is data mining and machine learning, with a focus on developing effective and efficient data analysis techniques for emerging big data intensive applications. He has published prolifically in refereed journals and conference proceedings, including IEEE TKDE, IEEE TMC, ACM TKDD, ACM SIGKDD, UCAI.

Hui Xiong (SM07) is currently a Full Professor and Vice Chair of the Management Science and Information Systems Department, and the Director of Rutgers Center for Information Assurance at the Rutgers, the State University of New Jersey, where he received a two-year early promotion tenure (2009), the Rutgers University Board of Trustees Research Fellowship for Scholarly Excellence (2009), and the ICDM-2011 Best Research Paper Award (2011). He received the B.E. degree from the University of Science and Technology of China (USTC), China, the M.S. degree from the National University of Singapore (NUS), Singapore, and the Ph.D. degree from the University of Minnesota (UMN), USA. His general area of research is data and knowledge engineering, with a focus on developing effective and efficient data analysis techniques for emerging big data intensive applications. He has published prolifically in refereed journals and conference proceedings (3 books, 60+ journal papers, and 90+ conference papers). He is a co-Editor-in-Chief of Encyclopedia of GIS, an Associate Editor of IEEE Transactions on Data and Knowledge Engineering (TKDE), IEEE Transactions on Big Data (TBD), ACM Transactions on Knowledge Discovery from Data (TKDD), and ACM Transactions on Management Information Systems (TMIS). He has served regularly on the organization and program committees of numerous conferences, including as a Program Co-Chair of the Industrial and Government Track for the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), a Program Co-Chair for the IEEE 2015 International Conference on Data Mining (ICDM), and a General Co-Chair for the IEEE 2013 International Conference on Data Mining (ICDM). He is an ACM Distinguished Scientist and a senior member of the IEEE.