International Journal of Web Services Research
Volume 15 • Issue 3 • July-September 2018

Cross-Urban Point-of-Interest Recommendation for Non-Natives

Tao Xu, Wuhan University, Wuhan, China
Yutao Ma, Wuhan University, Wuhan, China
Qian Wang, Wuhan University, Wuhan, China

ABSTRACT

This article describes how understanding human mobility behavior is of great significance for predicting a broad range of socioeconomic phenomena in contemporary society. Although many studies have been conducted to uncover behavioral patterns of intra-urban and inter-urban human mobility, a fundamental question remains unanswered: To what degree is human mobility behavior predictable in new cities—a person has never visited before? Location-based social networks with a large volume of check-in records provide an unprecedented opportunity to investigate cross-urban human mobility. The authors' empirical study on millions of records from Foursquare reveals the motives and behavioral patterns of non-natives in 59 cities across the United States. Inspired by the ideology of transfer learning, the authors also propose a machine learning model, which is designed based on the regularities that they found in this study, to predict cross-urban human whereabouts after non-natives move to new cities. The experimental results validate the effectiveness and efficiency of the proposed model, thus allowing us to predict and control activities driven by cross-urban human mobility, such as mobile recommendation, visual (personal) assistant, and epidemic prevention.

KEYWORDS

Cross-Urban Mobility, Group Profiling, Location-Based Service, Machine Learning, POI Recommendation

1. INTRODUCTION

Predicting human mobility is a fundamental question for a broad range of applications (Song et al., 2010), including urban planning (Yuan et al., 2012), epidemic forecasting (Dalziel et al., 2013), product advertising (Kirchner et al., 2012) and so on. It is vitally significant to understand socioeconomic phenomena embodying spatiality and human movement by unfolding human mobility patterns (Yan et al., 2013). Therefore, numerous researchers have attempted to uncover and model human mobility patterns (Gonzalez et al., 2008; Noulas et al., 2012; Hasan et al., 2013a; Schneider et al., 2013; Barchiesi et al., 2015; Pappalardo et al., 2015; Gallotti et al., 2016), to provide a deeper understanding of individual and collective mobility behaviors. In recent years, the rapid advances in mobile computing and social networking services empower people to share and use their locations and location-related content in location-based social networks (LBSNs) (Bao et al., 2015) such as Foursquare (https://foursquare.com) and Gowalla[1]. As a new type of data source containing extensive

DOI: 10.4018/IJWSR.2018070105

geo-tagged data with high position resolution and at large spatial scales, these LBSNs have enabled more exciting and critical studies on human mobility than those traditional ways of mobile phone and taxi (Noulas et al., 2012; Barchiesi et al., 2015; Hasan et al., 2013b; Wang & Stefanone, 2013; Wu et al., 2014; Zhang et al., 2015a; Zhang et al., 2015b; Huang & Wong, 2015; Hess et al., 2016).

In statistical physics and computer science, most of the previous studies conducted on large-scale check-in data from LBSNs focused mainly on intra-urban human mobility patterns and dynamics. In addition to the models proposed in statistical physics (Noulas et al., 2012; Hasan et al., 2013a; Schneider et al., 2013; Wu et al., 2014; Huang & Wong, 2015), machine learning based prediction models for points of interest (POIs) have attracted much attention, along with the increasing popularity of recommender systems (Hess et al., 2016). To the best of our knowledge, many types of information are considered as features when training such prediction models (Bao et al., 2015; Hess et al., 2016). These features include user's check-in history, content information (such as POI properties, user preferences, and sentiment indications) (Gao et al., 2015; Majid et al., 2013), the geographical influence based on the distribution of geographic distance (Ye et al., 2011; Liu et al., 2015), temporal influence (Yuan et al., 2013; Hosseini & Li, 2016), and social ties between friends (Ye et al., 2011; Zhou et al., 2015; Huang et al., 2015). In particular, a few elaborate hybrid models (Zhang et al., 2015b; Gao et al., 2015; Liu et al., 2015; Hosseini & Li, 2016; Huang et al., 2015; Yin et al., 2016), which utilize various combinations of the information mentioned above, have also been made to achieve higher recommendation accuracy for commercial purposes.

However, our research presented in this paper differs from those mentioned above in the areas of statistical physics and computer science. The primary goal of this work is to answer the intriguing question: To what degree individual human mobility behaviors are predictable in new cities? Due to the data sparsity problem, it is hard to predict human mobility behavior in a town an individual has never visited. Recent studies (Noulas et al., 2012; Gallotti et al., 2016; Liu et al., 2014; Jurdak et al., 2015; Liang et al., 2015; Zhao et al., 2016a) have discovered some universal laws on distance effect for human urban mobility across a variety of cities (also known as inter-city or inter-urban mobility), but the underlying patterns of human movements in new cities remain unclear. Also, we still lack prediction methods for human cross-urban whereabouts after people leave their cities of residence to a new one. Fortunately, machine learning provides robust methods and tools for understanding human behavior (Subrahmanian & Kumar, 2017). More importantly, the answer to this question would lead to valuable applications in emerging fields for non-native visitors, such as mobile recommendation (Yang et al., 2017), virtual (personal) assistant (Saad et al., 2017), and location-based advertising (Ketelaar et al., 2017).

In this work, we extend the research area of human mobility and first conduct an exploratory study on cross-urban human mobility patterns using the check-in data of a well-known LBSN. We carried out a large-scale empirical study on the dataset from Foursquare, including more than 50 thousand users and 3.5 million check-ins in 59 US cities. Although a few previous studies (Noulas et al., 2012; Gallotti et al., 2016; Liu et al., 2014; Jurdak et al., 2015; Liang et al., 2015; Zhao et al., 2016a) discussed the connection between spatial interactions and human mobility trends, our work highlights the motives and regularities of cross-urban human mobility behavior, regardless of the distance between cities. Our primary findings are enumerated as follows:

1. The regularity of changes in the categories of favorite POIs visited by non-natives (called visitors for short) within 24 hours a day is revealed to understand the motives behind their mobility behaviors better;
2. A few group profiles of different types of visitors and their noticeable check-in patterns are identified and then used for personal mobility behavior prediction.

Considering the promising results of transfer learning (or inductive transfer (Pan & Yang, 2010) in machine learning) in solving urban computing tasks (Wei et al., 2016), in this work, we also propose a heuristic model based on transfer learning to forecast visitors' cross-urban whereabouts in those cities they have never visited. Moreover, the model is verified to be valid by the evidence that its recommendation accuracy remains at relatively high levels when compared with baseline approaches.

The rest of this paper is structured as follows. Section 2 analyzes visitors' changing preferences for favorite POI categories over time, and Section 3 profiles a few representative groups of visitors according to their similar check-in patterns. Section 4 details our recommendation model, evaluation metric, experimental setups, empirical results, and possible application scenarios. Section 5 discusses an exciting by-product of our work and potential threats to the validity of the results. Section 6 introduces the previous works related to our study. Finally, Section 7 concludes this paper and presents our future work.

2. CHANGES IN VISITOR'S PREFERENCES FOR POPULAR POIS

2.1 Categorical-Temporal Distribution of POIs

A check-in record of a specific point location in LBSNs usually has two core elements, namely POI category and check-in time. Some companies like Google and Foursquare are generic and will collect and categorize POIs for any interest. For example, the category of the Newark Liberty International Airport, a well-known POI in the city of Newark, is *Airport*. Here, we use *Algorithm 1* to calculate the categorical-temporal distribution of POIs, an indicator for visitor's time-varying preferences for various types of POIs. Table 1 shows the symbols used in the algorithms of this paper.

In *Algorithm 1*, one day is divided into 24 time slots. The algorithm conducts a statistical analysis for check-in records of each user according to check-in time, and it generates a $24*|C|$ temporal-categorical matrix for each user in which the value of each cell $<i, j>$ indicates the frequency of a user's visits to j^{th} POI category during time slot i.

Table 1. Symbols used in the algorithms

Symbol	Description
u, U	u: visitor name; U: the set of visitors
c, C	c: POI category name; C: the set of POI categories
V	the set of visitors' check-in category vectors
V_u	the check-in category vector of visitor u
D	the collection of visitors' check-ins for POIs
D_u	the collection of check-ins of visitor u for POIs
Jm	the matrix of Jaccard similarity coefficient for every pair in V_u, and $Jm <i, j> = Jm <j, i>$
HCT	the hierarchical cluster tree of visitors
HCT_u	the node of visitor u in the hierarchical cluster tree
UC	the set of visitor clusters with similar check-in patterns
UC_j	a group of visitors with similar check-in patterns
$JSC <V_i, V_j>$	Jaccard similarity coefficient between V_i and V_j
tc	the temporal-categorical distribution of visitors
tc_u	the temporal-categorical distribution of visitor u

Algorithm 1. Temporal-categorical distribution calculation ($O(|D|)$)

Input: U, D
Output: tc
Create empty tc;
For each $u \in U$ **do**
    Create empty time vector T; //T_j represents a time slot, and $T_j \in [0:23]$
    For each $D_{u,i} \in D_u$ **do**
        Get t from $D_{u,i}$; //check-in time
        Get c from $D_{u,i}$; //POI category
        If T_j exists and $t \in [T_j\text{-}1, T_j\text{+}1]$ **do**
            $tc_u[c][(t + T_j)/2] \leftarrow tc_u[c][T_j] + 1$;
            Delete $tc_u[c][(t + T_j)/2]$;
        Else do
            $tc_u[c][(t + T_j)/2] \leftarrow 1$;
        End
    End
    $tc \leftarrow tc \cup tc_u$;
End

2.2. Mining the Top 30 Categories of Popular POIs

In Figure 1, the horizontal axis represents the 24 hours (indicated by the hours passed since midnight, from 0 to 23), while the vertical axis denotes the POI category associated with various POIs. Each circle in the 2-D coordinate system stands for the popularity of a POI category within a given period, which is measured by the number of check-ins and the number of visitors. Larger circles with a darker color denote those widely-watched POI categories, such as *Bar* (at “23”) and *Coffee Shop* (at “8”).

The top 30 POI categories depicted in Figure 1 fall within the scope of eight broad types. The first one is entertainment, including *Multiplex*, *General Entertainment*, *Other Great Outdoors*, *Sports Bar*, *Nightclub*, *Gay Bar*, *Café*, *Pub*, *Bar*, and *Coffee Shop*. The second one is diet, which includes *Diner*, *Fast Food Restaurant*, *Burger Joint*, *Pizza Place*, *Italian Restaurant*, *Sandwich Place*, *Mexican Restaurant*, and *American Restaurant*. The third one, called shopping, covers *Clothing Store*, *Mall*, *Department Store*, and *Grocery Store*. *Train Station*, *Subway*, and *Airport* belong to the fourth one known as transportation. *Gym* and *Gym/Fitness Center* fall within the scope of fitness, the fifth one. The remaining three types are accommodation, religious activity, and sightseeing, including *Hotel*, *Church*, and *Park*, respectively.

Due to the spatial difference between one’s place of residence and a new city, this categorization of popular POIs for visitors is a bit different from those for the intra-urban travel demands of natives (Wu et al., 2014; Bagrow & Lin, 2012). For example, a native may pay more attention to POIs near home and workplace, while a visitor might prefer venues of shopping and entertainment. Besides, these categories of popular POIs differ in both check-in time and check-in frequency. For instance, some visitors tend to check in POIs belonging to diet (e.g., *American Restaurant* and *Mexican Restaurant*) at meal times, whereas others prefer to visit places of entertainment (e.g., *Bar* and *Pub*) from sunset to midnight. This result indicates that the visitors’ cross-urban mobility activities are mostly affected by basic human needs (i.e., food, transportation, and accommodation) (Tay & Diener, 2011), their habits and customs (such as coffee drinking), and local recreational activities.

2.3. Overall Change in Popular POI Categories

For the most frequently visited POI categories, the change in check-in frequency over 24 hours a day is illustrated in Figure 3a. Circular-arc-shaped parts of the ring, painted in different colors, denote the hours passed since midnight, and the terms arranged outside each of the parts represent the top-*k* popular POI categories within a given period. Since a high number of check-in events occurred during the period spanning from 18 o’clock to 19 o’clock (i.e., the part labeled with “18”), we use a

Figure 1. The top 30 categories of popular POIs visited by non-natives in 24 hours

threshold to filter out less popular POI categories at different times. The value of this threshold is set to the number of check-ins for the 30th popular POI category within the hour (i.e., *Theater*). Also, each curved line between two POI categories within the ring indicates the corresponding semantic similarity between them, and we retain and display only those curved lines whose values of *Lin Similarity* (Lin, 1998) are greater than or equal to 0.8, e.g., *Mall* and *Department Store*. Figure 3b shows how to read the visualization of this map according to an example of the three hours from "19" to "21".

In general, the visitors' cross-urban check-in behaviors are significantly affected by human circadian rhythms. On the one hand, the visitors have more affection for the entertainment POIs (e.g., *Bar* and *Nightclub*) between the hours of midnight and 4 am, although this period of hours is considered to be bedtime for most people. On the other hand, their check-in behaviors present a different kind of diversity concerning POI categories and activity levels (measured by the number of POI categories) in the rest of a day. In the early morning, their check-ins focus mainly on fitness (e.g., *Gym* and *Gym/Fitness Center*), entertainment (e.g., *Coffee Shop* and *Café*), and transportation (e.g., *Subway* and *Airport*). As time goes on, religious activity (e.g., *Church*), business (e.g., *Tech Startup*) and shopping (e.g., *Grocery Store* and *Mall*) also become popular. Around noon diet (e.g., *Sandwich Place* and *American Restaurant*) is the preferred choice for the majority of the visitors. In the afternoon, the activity level of their check-in behaviors, however, decreases until dinner time. Then, the non-natives start to enjoy their nightlife, including bars, restaurants, pubs, nightclubs, and so on, late into the night. The activity level of their check-in behaviors, unsurprisingly, continues to fall as it is getting late.

Besides, Figure 3a also conveys the meaning of a few common behaviors which are related closely to the American way of life, customs, and culture. Firstly, the frequent occurrence of *Airport*, *Hotel*, and *Subway* suggests that these visitors, indeed, like to travel by plane, stay in hotels, and go anywhere by subway. This result further demonstrates that it is necessary to meet the core human needs (i.e., food, transportation, and accommodation) first. Secondly, coffee and fitness have become the quintessential symbol of the American way of life (Lichtenstein et al., 2006; Mozaffarian et al., 2011), and even in different cities, most of the visitors tend to maintain such an active lifestyle, regardless of the distance from their cities of residence. This result is well-supported by the frequency of occurrence of *Coffee Shop* and *Gym/Fitness Center*. Thirdly, according to the evidence similar to the ones listed above, we deduce that the visitors' leisure and entertainment activities at night are affected by local bar culture and indoor shopping. It is worth noting that indoor shopping has been woven tightly into American culture since the early 1980s (Jackson, 1996; Palan et al., 2010) and is a favorite way of relaxation and entertainment, especially when people intend to purchase local products.

3. GROUP PROFILING BASED ON SIMILAR CHECK-IN BEHAVIORS

3.1. Hierarchical Clustering Algorithm

Nowadays, user profiling is a commonly-used technique for Internet companies to capture the explicit digital representation of a person's identity. Compared with user profiles, a group profile characterizes specific behaviors or other features of a class of people. To identify group profiles of different visitors, we use a hierarchical clustering algorithm of machine learning to categorize those visitors with similar check-in patterns. Our algorithm is designed based on the Jaccard Index (Jaccard, 1901) and Hierarchy clustering schemes (Johnson, 1967), and the details of the algorithm written in pseudo-code refer to *Algorithm 2*.

Algorithm 2 is a hierarchical clustering algorithm based on the Jaccard index. It first generates $|U|$ initialized nodes, each of which has a vector of the POI categories that the user has ever visited. In each iteration, it chooses the most similar two users with the maximum value of the Jaccard index

Algorithm 2. Jaccard index-based Hierarchical Clustering (O(log |*U*|)

```
Input: V, U, Jm
Output: UC, HCT
Initialize HCT containing |V| leaf nodes, each of which consists of V_u;
For each V_i ∈ V, i ∈ U do
    For each V_j ∈ V − {V_i}, j ∈ U do
        Compute JSC <V_i, V_j> between V_i and V_j;
        If JSC <V_i, V_j> > 0 do
            Jm <i, j> ← JSC <V_i, V_j>;
        End
    End
End
While Jm is not empty do
    Find the biggest Jm <i, j>;
    U ← (U ∪ {new}) − {i, j}; //new is a new node composed of the two most similar users
    V_new ← V_i ∩ V_j;
    Create node HCT_new containing V_new;
    Set node HCT_i and HCT_j to be the children of HCT_new, then add HCT_new to HCT;
    For each k ∈ U do
        Delete Jm <i, k> and Jm <j, k>;
        Jm <new, k> ← JSC <V_new, V_k>;
    End
End
UC = {all leaf nodes of HCT_i | |V_i| of HCT_i is greater than five};
```

to form a new intermediate node; then, it adds the intermediate node to the set of the remaining nodes after deleting the selected nodes. This procedure ends until the set of all the remaining nodes contains less than five POI categories. As a result, it generates several clusters representing different groups of users who have checked in at least five POI categories.

3.2. Mining Representative Groups of Visitors

According to similar check-in behaviors, visitors from different cities can gather into representative groups (see the right side of Figure 2) using our clustering algorithm. The left side of Figure 2 is a heat map showing how the visitors checked in POIs, where each row is a vector of a person's check-ins, and each column is a POI category. In the heat map, blue and white points denote that a visitor has ever visited and never visited POIs of a POI category, respectively. The hierarchical clustering tree, which is in the middle of Figure 2, is obtained by using *Algorithm 2*. Owing to space limitations, we only present the top ten representative groups as well as their group profiles on the right side of Figure 2. Each black rectangular box that contains several small squares in different color denotes a group of visitors, and the legends for these little squares locate in the lower right corner of Figure 2. On the right side of each box is a 2-D shadow map, showing how active these groups of visitors are and how much their check-in activities vary over time. Here, the horizontal axis represents the 24 hours, while the vertical axis represents the number of check-ins for POIs that belong to the POI categories within the corresponding box within a given period.

Here, we show the top ten groups of visitors (denoted by G_i, $i \in [1, 10]$). G_1 prefers to visit POIs belonging to *Grocery Store*, *Department Store*, *Mall*, *Fast Food Restaurant*, and *Bank* in the afternoon and evening. Besides shopping places (i.e., *Grocery Store*, *Department Store*, and *Mall*), G_2 also often checks in the POIs that belong to *Bar*, *Coffee Shop*, and *Mexican Restaurant*, which indicates that the visitors in this collection would like to enjoy some leisure time as well. Interestingly, G_3 shows a distinct preference for the *Gas Station/Garage* POIs, possibly suggesting that self-driving tour is a favorite way for the group members.

G_4 and G_6 resemble G_1 regarding check-in behavior, but a significant difference between them lies in the POI categories of shopping and diet. Besides *Department Store* and *Grocery Store*, G_1 would like to visit the *Mall* POIs. Unlike the preference of G_1 for *Fast Food Restaurant*, G_4 and G_6 share a common interest in *American Restaurant* and *Pizza Place*, and they enjoy *Mexican Restaurant* and

Figure 2. Profiling the top 10 representative groups of visitors

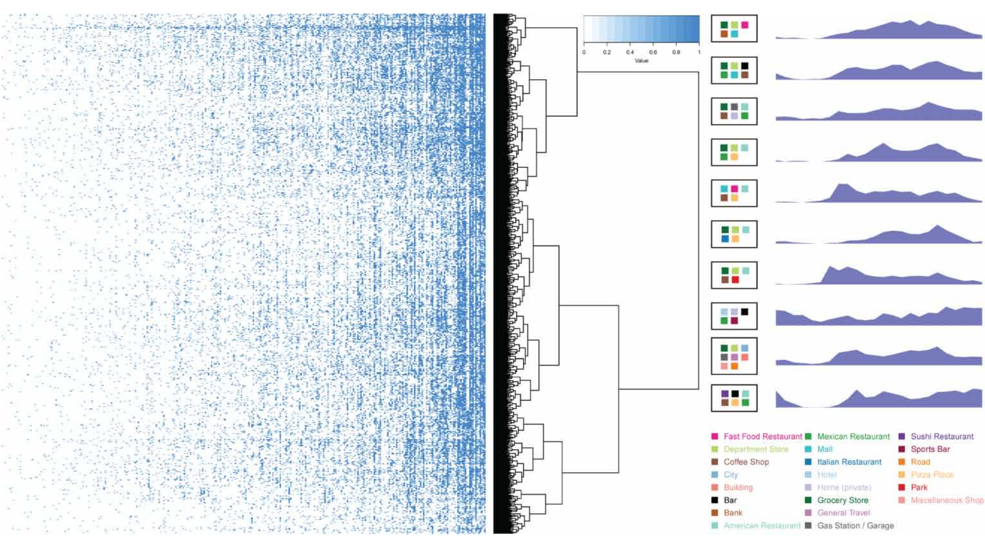
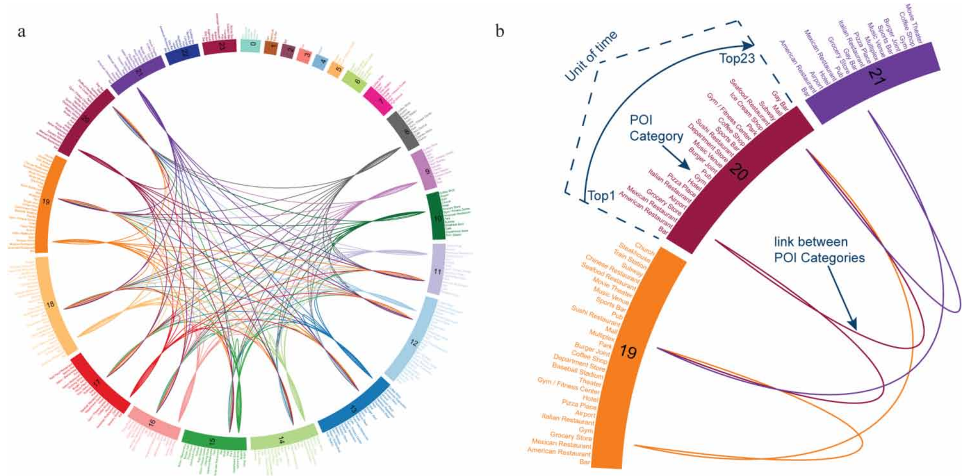


Figure 3. A map of popular POI categories that vary over time



Italian Restaurant, respectively. Moreover, G_5, G_7, and G_{10} also exhibit similar check-in behavior, although they have their particular inclinations for food (e.g., *Sushi Restaurant* for G_{10}), shopping (e.g., *Mall* for G_5), and recreation (e.g., *Park* for G_7). Compared with G_3, G_9 with a more complex behavior tends to seek for the beautiful roadside sceneries across the United States, such as *Building*, *Road*, and *City*. Finally, unlike the groups mentioned above, G_8 is more likely to be composed of energetic young people because they spend more time visiting those POIs belonging to *Bar*, *Sports Bar*, and *Home (private)* for individuals, from late night to early morning.

4. RECOMMENDING POIS FOR VISITORS IN NEW CITIES

4.1. Recommendation Model

Inspired by the ideology of transfer learning, we design a basic recommendation model and the model's three variants that incorporate diverse information such as semantic similarity and group profile. These models take user ID, city name, and time as input, and output the top-*k* POIs for a given visitor. *Algorithm 3* describes the procedures of the four kinds of recommendation models which utilize different types of information.

Assuming that different people have their mobility preferences, the primary temporal-categorical model (*BTC*), which is built based on the personal distribution of POI categories over time for each visitor in the train set, is used as a baseline method in this study. The categorical-temporal distribution of POIs (also called prior knowledge) is the outcome of *Algorithm 1*, which takes into consideration the following features: user, the longitude and latitude of a POI, POI category, and time. The details of the *BTC* model refer to *Case 1* in *Algorithm 3*.

Compared with *BTC*, the *BTCLS* model incorporates the *Lin Similarity* (Lin, 1998) between various POI categories. Despite the fact that many POI categories have different names, there exist semantic similarities between them. For example, the semantic resemblance between *American Restaurant* and *Mexican Restaurant* is evident. The details of the *BTCLS* model refer to *Case 2* in *Algorithm 3*.

The *BTCCS* model, which depends on the assumption that visitors who have a similar check-in pattern are likely to share similar behavioral preference, uses *Algorithm 2* to group visitors with similar check-in behaviors. The recommendation process of this model resembles that of user-

Algorithm 3. POI recommendation (O(|*U*|*|*T*|))

```
Input: U, C, D, UC
Output: Top-k recommended POIs
Initialize tc of visitors calling Algorithm 1;
For each u ∈ U do
    Get time vector T of visitor u from test data extracted from D;
    For each T_j ∈ T do
        Procedure SR:
            Create an empty recommended category vector VC;
            For each c ∈ C do
                If tc_u[c][t] > 0 and t ∈ [T_j-1,T_j+1] do
                    VC ← VC ∪ c;
                    Break;
                End
            End
        End
        Case 1://BTC model
            Break;
        Case 2://BTCLS model, SimLin() is a function to compute the Lin Similarity
            For each VC_j ∈ VC do
                SimVC_j ← {c | c ∈ C and SimLin(c, VC_j) > 0.5};
            End
            VC ← VC ∪ (∪_{j < |VC|} SimVC_j);
            Break;
        Case 3: //BTCCS model
            Find UC_u that visitor u belongs to;
            For each visitor in UC_u, execute Procedure SR, then get |UC_u| VC;
            VC ← ∪ (VC_u of each visitor of UC_u);
            Break;
        Case 4: //HYBRID model
            For each VC_j ∈ VC do
                SimVC_j ← {c | c ∈ C and SimLin(c, VC_j) > 0.5};
            End
            VC ← VC ∪ (∪_{j < |VC|} SimVC_j);
            Find UC_u that visitor u belongs to;
            For each visitor in UC_u, execute Procedure SR, then get |UC_u| VC;
            VC ← ∪ (VC_u of each visitor of UC_u);
            Break;
    End
    Sort VC according to the popularity of POI categories;
    Find the top three popular POIs of each category in VC;
    Recommend the top k POIs to u at time T_j;
End
```

based collaborative filtering (Wang et al., 2006) in machine learning, and its details refer to *Case 3* in *Algorithm 3*.

The *HYBRID* model is a hybrid model that incorporates the categorical-temporal distribution, *Lin similarity* between POI categories, and group profiles. The procedure of the *HYBRID* model, please refer to *Case 4* in *Algorithm 3*.

4.2. Baseline Methods

Collaborative filtering (CF) is a commonly-used technique in recommender systems. Memory-based CF uses user rating data to calculate the similarity between users or items and includes two typical methods, namely user-based and item-based CF. In this study, we employ user-based CF as a baseline method. This baseline method searches for users who share the same (or similar) check-in patterns with the target user and then uses the POIs from the top *k* like-minded users found in the previous step to predict possible POIs for the target user (see Algorithm 3).

Matrix factorization (MF), also called matrix decomposition, is a factorization of a matrix into a product of matrices. Some of MF methods such as singular-value decomposition (SVD) have also been used in recommender systems. In this study, we employ the SVD method of MF as a baseline method, which falls under the category of item-based CF. This baseline method finds the similarity between all pairs of POIs and then uses the most similar POIs to the target user's already-visited POIs to generate recommendations.

Recently, Zhang & Wang (2016) proposed a POI recommendation approach based on cross-region collaborative filtering (CRCF) to recommend POIs for users who travel to a new city or region that they have never visited before. CRCF combines the predicted rating on the content of a POI and the predicted rating on the location of the POI. Content refers to any descriptive information about a POI and its function, such as user-generated tags and comments. Considering the relevance of CRCF with our work, we also employ it as a baseline method.

4.3. Evaluation Metric

A metric *Accuracy* is used to measure the prediction accuracy for visitor *u*, defined as below:

$$Accuracy_u = \frac{\sum_{i=1}^{|M|} T_i^k}{|M|} \tag{1}$$

where *M* is the set of cities the visitor visited (except the city of residence), |*M*| is the cardinality of the set *M*, and T_i^k is defined as follows:

$$T_i^k = \begin{cases} 1, & if\ P_k \cap R \neq \varnothing \\ 0, & otherwise \end{cases} \tag{2}$$

Here, P_k is the set of the top-*k* POIs recommended by a given model and *R* is the set of all the POIs visited by the visitor in the test set. The average accuracy calculated for all the visitors in question is used to evaluate the performance of the four recommendation models:

$$AvgAcc = \frac{\sum_{i=1}^{|U|} Accuracy_i}{|U|} \tag{3}$$

where *U* is the set of visitors for evaluation and |*U*| is the cardinality of *U*.

4.4. Experimental Setups

4.4.1. Data Collection

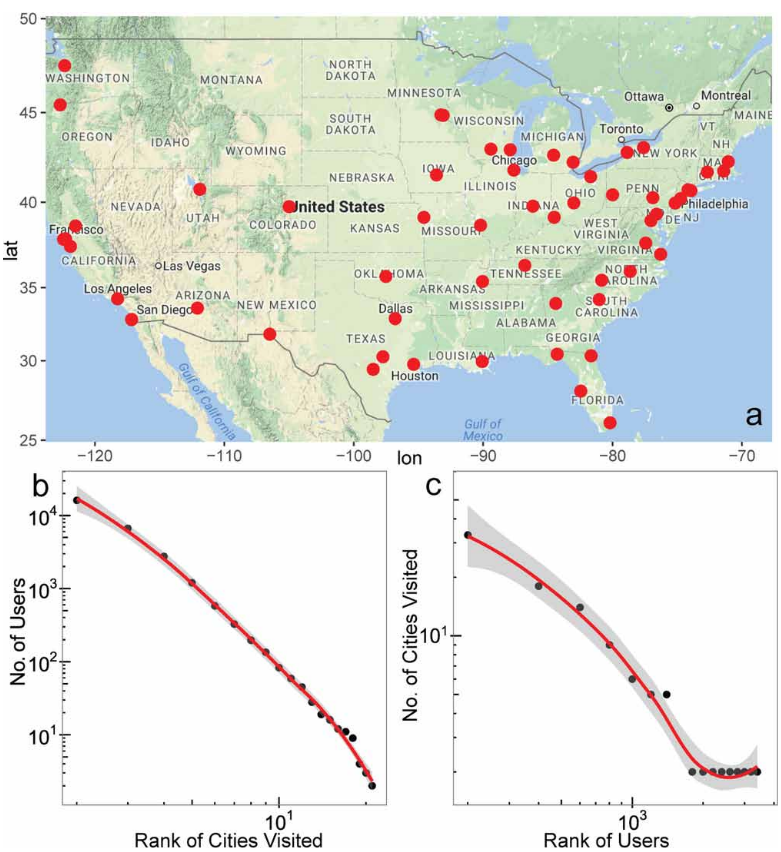
Foursquare is one of the most active LBSNs, and its primary function enables users to share their locations with their friends. The original dataset used for our research, which is ready for download at the website (https://sites.google.com/site/yangdingqi/home/foursquare-dataset), contains 18-month global-scale check-in data from Foursquare, including 33,278,683 check-ins by 266,909 users on 3,680,126 venues (in 415 cities across 77 countries). The 415 cities are the most popular cities visited by Foursquare users in the world, each of which contains at least 10,000 check-ins (Yang et al., 2015, 2016). We retained 3,555,267 check-in records by 50,801 Foursquare users in 59 US cities. Figure 4a presents the spatial distribution of the 59 cities across the United States using Google Maps. This dataset includes 501,900 POIs and 427 POI categories.

In this study, the dataset used in our experiments excluded those users whose check-in records exist in only one city (also known as natives). After analyzing each user's city of residence under the assumption that everyone would leave the maximum check-in records in the city where he or she lives, we identified 16,141 visitors out of the 50,801 users. Figure 4b depicts the popularity of the 59 US cities for visitors. Each point (with values of x and y) in the 2-D coordinates denotes that there are x cities which attract at least y visitors. The distribution of the points follows a power law, according to the best fitting curve at a 99% confidence interval (grey shadow). This result indicates that the minority of the cities (such as Los Angeles and New York) attract the most attention. Figure 4**c** shows the 16,141 visitors' preferences for the US cities. Each point represents that there are x visitors who have visited at least y cities. Similarly, the distribution of the points roughly follows a power law, suggesting that most of the visitors have visited only several cities across the United States.

4.4.2. Experimental Environment

We conducted all of our experiments on a DELL Precision Tower 5810 (GPU workstation) running 64-bit Windows 10. The GPU workstation has an Intel Xeon Processor E5-1620 v3 (4 cores, 3.5GHz), 16GB RAM, and a GPU of NVIDIA GeForce GTX 960. All the POI recommendation models and algorithms were implemented by using C++11 in Microsoft Visual Studio 2013, with the Boost C++ Libraries (Version 1.60.0) supported.

Figure 4. Statistics of the check-in data used in this study



4.5. Empirical Results

4.5.1. Prediction Performance for the Entire Set of Visitors

For each of the 16,141 visitors under discussion, all the check-in records in his/her city of residence were selected as training data, and those check-in records in the other cities the visitor visited (except the one used for training) were used as test data. Note that the training set and test set contain about 143 thousand records and 42 thousand records, respectively. Unfortunately, all the baseline approaches, namely CF, MF, and CRCF, failed to output any result when our models accomplished the task of POI recommendation for all the visitors in question.

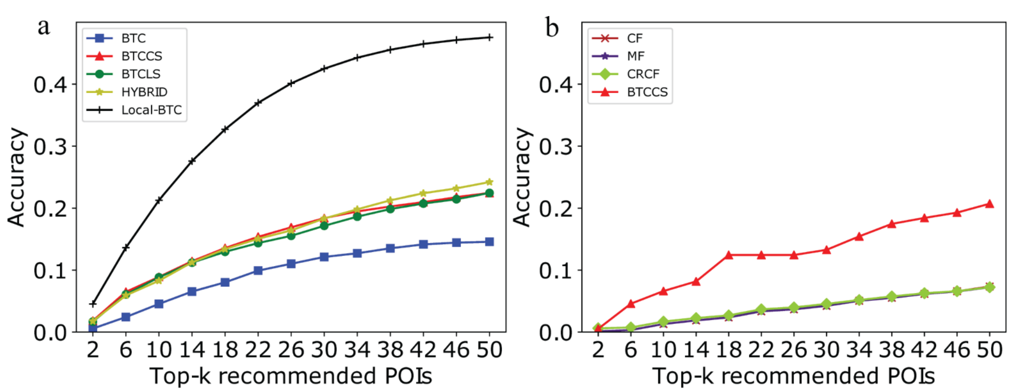
As shown in Figure 5a, the vertical axis denotes the value of *AvgAcc*, and the horizontal axis means the number of recommended POIs. Note that the model *Local-BTC* is a *BTC* model which is trained by visitors' local check-in records in their cities of residence and acts as a baseline method of intra-urban POI prediction. In general, *BTCCS*, *BTCLS*, and *HYBRID* can obtain approximate results, and they are better than the model *BTC*. It is noteworthy that the best accuracy we obtained by using *BTCCS*, *BTCLS*, and *HYBRID* is, on average, half that of *Local-BTC*, especially when recommending a small number of POIs. Because it is impossible to collect individual check-in records in those new cities that people have never visited before (Bao et al., 2015; Zhao et al., 2016b), this would no doubt contribute to addressing the data sparsity problem (that results in the cold start problem (Schein et al. 2002)) in POI recommender systems. Therefore, the results presented in Figure 5a are helpful to answer the research question of this paper; that is to say, visitors' cross-urban mobility behaviors are, to some extent, predictable from their historical check-in records in given cities of residence, according to some simple, reproducible behavioral patterns discovered and described above.

4.5.2. Prediction Performance for the 100 Most Active Visitors From the Total Visitors

To make a comparison between our method and the baseline approaches on a small-scale dataset, we downsampled the 100 most active users to form a subset of the total visitors and extracted their check-ins in different cities from the original dataset introduced in Subsubsection 4.4.1. This small-scale dataset contains the top 100 visitors, 26,540 POIs, and 17,824 check-in records. Note that visitor's check-ins are grouped by cities. For each visitor in the small-scale dataset, the selection of training data and test data has the same settings as described in Subsubsection 4.5.1.

Although CRCF performs slightly better than CF and MF, all the three baseline methods produce similar results, but not good enough. One of the main reasons for this is the data sparsity problem. It is evident from Figure 5b that *BTCCS* outperforms the three baseline methods regarding the average accuracy except when *k* equals one or two. Moreover, the performance of *BTCCS* on this small-scale dataset is approximate to that achieved on the entire set of visitors, suggesting good scalability of

Figure 5. A comparison between our models and the baseline methods



this model. We argue that the scalability of our models arises from the general motives and universal behavioral patterns of visitors found in this study. In particular, these findings are roughly less affected by the distance between cities and transformation between urban spaces.

4.5.3. Prediction Performance for Niche POIs

However, the diversity of human movements makes the absolute predictability in human mobility more challenging (Song et al., 2010). To test the capability of our models for predicting niche POIs (excluding those of the categories illustrated in Figure 1), also known as personalized POI recommendation, we conducted an additional experiment on four famous US cities, namely New York, Los Angeles, Washington D.C., and Chicago. If a POI category, which includes several POIs, has less than 500 check-in records, it is a niche category in the experiment. As depicted in Figure 6, areas colored in red in a density map have high densities, and the bottom 30 POI categories, with darker blue indicating lower popularities, are placed next to the density map. For each of the four cities, the distribution of visitors' check-in frequencies for POIs of niche categories scattered across the city is, of course, heterogeneous. This finding is compatible with the latest result reported by Yang *et al*. (2017). Even so, our models, especially *BTCLS*, also provide the adequate predictive power of personalized POI recommendation for visitors (i.e., about one-third that of *Local-BTC*), according to the result indicated in Table 2. Besides, they show a distinct advantage over the baseline methods in accuracy. The five models and baseline methods were trained based on all the check-in records in each visitor's city of residence and tested by those check-in records of niche POI categories in each of the four cities. Because niche POIs receive far less attention than popular ones, the value of top-k was set to 50 in this experiment. In brief, this result can, indeed, demonstrate certain predictability of cross-urban human mobility behavior by using our models.

4.6. Potential Application Scenarios

Our work could facilitate improved modeling and prediction of human movement in new cities. For instance, the core of our business is the recommendation model, which can be used for the scenarios of

Figure 6. Density maps of visitors' check-in frequencies for niche POIs in the four US cities

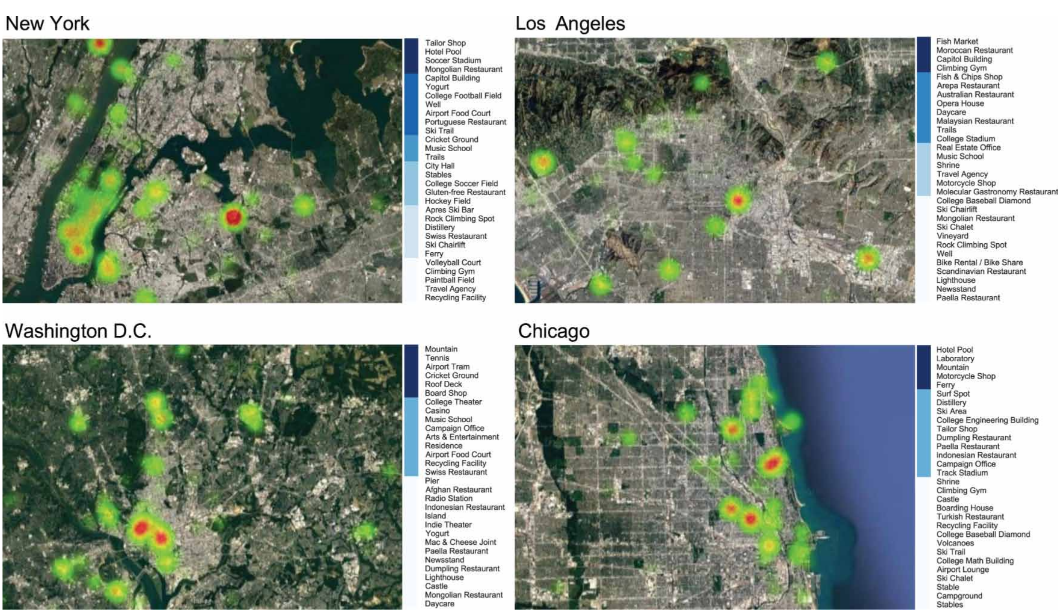


Table 2. Performance comparison of our models and the baseline methods for niche POIs over the four US cities

	Washington D.C.	Los Angeles	New York	Chicago	Average
BTC	0.075	0.117	0.092	0.123	0.102
BTCCS	0.131	0.117	0.125	0.167	0.135
BTCLS	0.234	0.091	0.180	0.196	0.175
HYBRID	0.131	0.078	0.132	0.190	0.133
CF	0.022	0.019	0.021	0.012	0.019
MF	0.024	0.017	0.022	0.018	0.020
CRCF	0.023	0.009	0.026	0.021	0.020
Local-BTC	0.630	0.599	0.562	0.457	0.562

mobile recommendation and location-based advertising to achieve the purpose of precision marketing for visitors. Here is a simple example of an intelligence app developed based on our work. If the app detects a visitor's real-time location in a new city, it will proactively recommend the visitor the most likely POIs according to the current time, the distances from these POIs to the visitor, and group profiles. Besides, our work lays a foundation for visual (personal) assistants. A smart visual assistant learns individual movement profile in his/her city of residence automatically. Before the owner of the app goes to another (new) city, the assistant will make suggestions on some possible POIs and the related trip scheduling.

Compared with traditional coarse-grained census data, LBSNs provide a vast amount of spatiotemporal data with high position resolution for more realistic mobility assumptions for human migration and population projection (Jurdak et al., 2015). Another promising implication of our work is for further research in human geography and demography. The combination of inter-urban human mobility studies and our work could contribute to a more comprehensive understanding of the underlying drivers for people moving at different spatial scales, such as distance effect, circadian rhythms, basic human needs, and social and cultural influence. More importantly, these mobility patterns and dynamics can apply to some essential activities driven by human mobility, such as risk early-warning for public emergency and epidemic prevention for visitors from various places. For instance, if some people infected with an infectious disease fly to a given city, epidemiologists can narrow down the range of possible POIs in the city, according to our models, to stop the spread of the disease as soon as possible.

5. DISCUSSION

5.1. Distance Effect

Our work reveals underlying patterns of cross-urban human mobility in the temporal dimension, but the related behavioral patterns of non-natives in the spatial dimension, e.g., location sequence and displacement distance, are complicated and need to be explored. Due to the lack of adequately labeled data for visitors' places of residence (e.g., *Hotel* POIs), it is hard to choose a reasonable starting point for each visitor in a new city. Therefore, for each of the 59 cities, we extracted and analyzed visitors' movement trajectories composed of a series of check-ins that occur within the same town. In this work, we employed the metric *distance*, which is calculated based on the geographic distance between two intra-city POIs corresponding to a pair of consecutive check-ins, to measure the mobility of individuals in a new city.

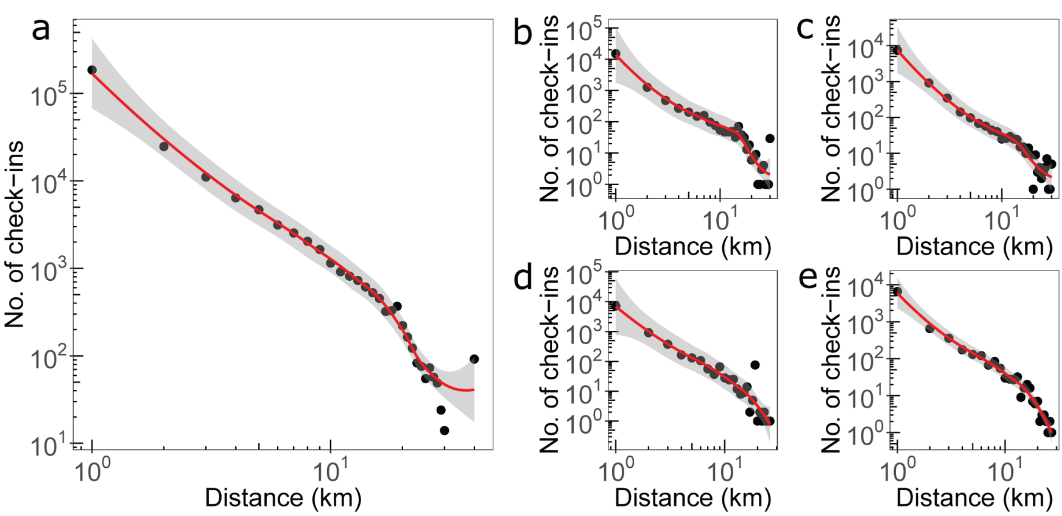
As shown in Figure 7, each point (with values of *x* and *y*) in the 2-D coordinates denotes that there are *y* check-ins that take place within the distance of *x* kilometers from the current position (or called the displacement distance *x* (Noulas et al., 2012)). Considering the time interval between any two consecutive check-ins in a trip may range from a few minutes to several days, we filtered out those noisy check-ins spanning over one week. The distributions of points in these five subplots in Figure 7 follow a power law with an exponential cutoff, according to the best fitting curve at a 99% confidence interval (grey shadow). Note the change in slope at around 15km.

The cumulative distributions of the number of check-ins vs. *distance* exhibit similar trends at different scales, although their shapes vary from city to city. As the value of *distance* increases, the number of check-ins decays in a universal power-law form with an exponential cutoff at the levels of both city and country. That is to say, despite long distance moves (*distance* > 30km), most of the visitors prefer nearby POIs after departing from their current locations. This result presents the first large-scale empirical evidence in the distance decay effect (Nekola & White, 1999) on non-native visitors. Besides, Figure 7 suggests that around 15km is the typical human radius of “jump” (or called “displacement” (Noulas et al., 2012)) in new cities as it takes about 15 to 30 minutes to drive such distance (by bus, taxi, or subway). Obviously, this distance is far smaller than 100km, the typical human radius of “reach” traveled from home (Mok et al., 2010; Cho et al., 2011), mainly due to different trip purposes and functional zoning (Dubrova et al., 2015) in various cities. Hence, the factor (distance) may not be a critical (or even an appropriate) variable to model universal patterns of human movements, which is consistent with the results of a few previous studies (Song et al., 2010; Noulas et al., 2012; Gallotti et al., 2016; Wu et al., 2014; Liu et al., 2014; Liang et al., 2015; Zhao et al., 2016a).

5.2. Threats to Validity

First, much recent literature has provided sufficient evidence that LBSNs such as Twitter can indeed be a useful proxy for tracking and predicting human movement (Hasan et al., 2013b; Wang & Stefanone, 2013; Wu et al., 2014; Huang & Wong, 2015; Liu et al., 2014; Jurdak et al., 2015; Yang et al., 2017), but the external validity of our results on other different LBSNs is yet to be tested. Second, there are two concerns about internal validity. On the one hand, the selection of 100 visitors and four famous US cities in our study may cause sampling bias. On the other hand, in our experiment, we only used three

Figure 7. The number of check-ins as a function of geographic distance traveled from the current position: (a) All the 59 US cities; (b) New York; (c) Los Angeles; (d) Washington D.C.; and (e) Chicago



baseline methods, implying that we may neglect some latest POI recommendation approaches. Third, as mentioned in some previous works (Song et al., 2010; Yan et al., 2013; Liu et al., 2014), individual heterogeneity (Gonzalez et al., 2008) requires the variability in mobility prediction at the individual level. It is worth noting that our proposed model, as a preliminary attempt to investigate cross-urban human mobility, also possesses the potential of more accurate personalized recommendation with sophisticated machine learning techniques such as deep learning. Besides, when training new prediction models, we will take into consideration multi-modal information about LBSNs, such as geographical influence, sequential influence (Zhang et al., 2014), and the indigenization coefficient between natives and visitors (Yang et al., 2017).

6. RELATED WORK

As mentioned above, predicting human mobility is an important scientific question across multiple research areas, such as statistical physics, computer science, and human geography. LBSNs have been recognized as a convenient proxy for predicting human movement, especially for the young. According to the recent surveys on this issue (Bao et al., 2015; Hess et al., 2016; Zhao et al., 2016b), previous studies paid more attention to intra-urban and inter-urban human movement in LBSNs. In this section, we present an overview of the previous works from the two perspectives.

6.1. Intra-Urban POI Recommendation

Unlike those traditional product recommendation systems, POI recommendation must take into account physical constraints and geographical influence, because the check-in behavior is affected by locations' geographical features. In computer science, the POI recommendation models for intra-urban trips are, in general, built by using machine learning techniques. Besides, there are four main influential factors used in POI recommendations: geographical influence, temporal dynamics, social relations, and content indications (Bao et al., 2015; Hess et al., 2016; Zhao et al., 2016b).

The general task of POI recommendation is to recommend the top-k POIs for users, resembling the recommendation task in traditional recommendation systems for books or movies. Researchers have proposed a range of models to incorporate different types of influential factors to accomplish this task. For example, Ye *et al.* (2011) proposed a fused model that combined the traditional user-based collaborative filtering method and two types of influential factors, namely geographical influence and social influence. The experimental results from two real-world datasets (i.e., Foursquare and Whrrl[2]) show that the proposed model performs better than several alternative recommendation approaches. Under a unified POI recommendation framework, Gao *et al.* (2015) modeled three types of content information (i.e., POI properties, user interests, and sentiment indications) jointly using matrix factorization. Unlike the fused model, such a joint model learns several influential factors together and then recommends POIs for users by the jointly learned model. Therefore, the experimental results from the Foursquare dataset indicate the significance of content information in explaining user behavior.

Successive POI recommendation, an extension of the general task of POI recommendation, is a specific POI recommendation task. It offers a given user personalized recommendations sensitive to the user's recent check-in rather than a general list of POIs (Cheng et al., 2013). For example, Zhao *et al.* (2016c) proposed the STELLAR model to provide time-aware successive POI recommendations. In particular, the model explicitly models the interactions among user, POI, and time using a ranking-based pairwise tensor factorization framework. The empirical results from the Foursquare and Gowalla datasets show that the STELLAR model outperforms all the baseline methods.

6.2. Inter-Urban Human Mobility Prediction

Human mobility across a variety of cities, known as inter-city or inter-urban mobility, has also attracted much attention from many research fields. Recent studies (Noulas et al., 2012; Gallotti et al., 2016; Liu et al., 2014; Jurdak et al., 2015; Liang et al., 2015; Zhao et al., 2016a) discovered some universal laws of the distance effect on inter-urban human mobility. For example, Noulas *et al.* (2012) discovered that the probability of moving from one place to another was inversely proportional to the total number of intervening opportunities between them. By fitting the gravity model, Liu *et al.* (2014) found that the observed spatial interactions (i.e., inter-urban movements) were ruled by a power-law distance decay effect. Liang *et al.* (2015) uncovered the probability that a trip would occur was inversely proportional to the size of the population located inside a circle of radius (which equals the travel distance) centered at the trip origin. However, the underlying patterns and motivations of human movements in new cities (also known as cross-urban human mobility) remain unclear.

7. CONCLUSION AND FUTURE WORK

In network science, individual human mobility becomes an emergent research field that sets out to find patterns and regularities governing human movements. In recent years, large-scale LBSN data sets available on human movements among a set of POIs have provided an unprecedented opportunity to investigate such a fundamental issue. However, most of the previous studies on POI recommendation are confined to intra-urban human mobility. Convenient transport facilitates long-distance, inter-urban individual movements. In this study, we explore the motives and behavioral patterns of non-natives in 59 cities across the United States using millions of records from Foursquare. Besides, we propose a machine learning model based on the findings of this study to recommend possible POIs for visitors when they come to a new city. The experimental results on different sizes of visitor groups indicate the effectiveness and efficiency of our method over two typical collaborative filtering techniques and one recently-proposed approach related intimately to our method.

The primary goal of our future work is to improve the personalized prediction performance of the proposed model by considering more types of influential factors, such as sequential influence and content information, as well as by using deep learning.

ACKNOWLEDGMENT

The authors thank Dr. Qiongyu Guo and Dr. Mudabbir Badar for their useful discussions on this paper. This work was supported by the National Basic Research Program of China (973 Program) under Grant No. 2014CB340404, the National Key Research and Development Program of China under Grant No. 2016YFB0800400, the National Natural Science Foundation of China under Grant Nos. 61672387, 61702378, and 61373167, the Wuhan Yellow Crane Talents Program for Modern Services Industry, the Strategic Team-Building Project of Scientific and Technological Innovation in Hubei Province, and the Fundamental Research Funds for the Central Universities under Grant No. 2042016kf0137.

REFERENCES

Bagrow, J. P., & Lin, Y.-R. (2012). Mesoscopic Structure and Social Aspects of Human Mobility. *PLoS One*, *7*(5), e37676. doi:10.1371/journal.pone.0037676 PMID:22701529

Bao, J., Zheng, Y., Wilkie, D., & Mokbel, M. F. (2015). Recommendations in location-based social networks: A survey. *GeoInformatica*, *19*(3), 525–565. doi:10.1007/s10707-014-0220-8

Barchiesi, D., Preis, T., Bishop, S., & Moat, H. S. (2015). Modelling human mobility patterns using photographic data shared online. *Royal Society Open Science*, *2*(8), 150046. doi:10.1098/rsos.150046 PMID:26361545

Cheng, C., Yang, H., Lyu, M. R., & King, I. (2013). Where You Like to Go Next: Successive Point-of-Interest Recommendation. In *Proceedings of the 23rd International Joint Conference on Artificial Intelligence* (pp. 2605–2611).

Cho, E., Myers, S. A., & Leskovec, J. (2011). Friendship and mobility: user movement in location-based social networks. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1082–1090). doi:10.1145/2020408.2020579

Dalziel, B. D., Pourbohloul, B., & Ellner, S. P. (2013). Human mobility patterns predict divergent epidemic dynamics among cities. In *Proceedings of the Royal Society of London B: Biological Sciences, 280*(1766). doi:10.1098/rspb.2013.0763

Dubrova, S. V., Podlipskiy, I. I., Kurilenko, V. V., & Siabato, W. (2015). Functional city zoning. Environmental assessment of eco-geological substance migration flows. *Environmental Pollution*, *197*, 165–172. doi:10.1016/j.envpol.2014.12.013 PMID:25541071

Gallotti, R., Bazzani, A., Rambaldi, S., & Barthelemy, M. (2016). A stochastic model of randomly accelerated walkers for human mobility. *Nature Communications*, *7*, 12600. doi:10.1038/ncomms12600 PMID:27573984

Gao, H., Tang, J., Hu, X., & Liu, H. (2015). Content-Aware Point of Interest Recommendation on Location-Based Social Networks. In *Proceedings of the 29th AAAI Conference on Artificial Intelligence* (pp. 1721–1727).

Gonzalez, M. C., Hidalgo, C. A., & Barabási, A.-L. (2008). Understanding individual human mobility patterns. *Nature*, *453*(7196), 779–782. doi:10.1038/nature06958 PMID:18528393

Hasan, S., Schneider, C. M., Ukkusuri, S. V., & González, M. C. (2013a). Spatiotemporal patterns of urban human mobility. *Journal of Statistical Physics*, *151*(1), 304–318. doi:10.1007/s10955-012-0645-0

Hasan, S., Zhan, X., & Ukkusuri, S. V. (2013b). Understanding urban human activity and mobility patterns using large-scale location-based data from online social media. In *Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing* (p. 6). doi:10.1145/2505821.2505823

Hess, A., Hummel, K. A., Gansterer, W. N., & Haring, G. (2016). Data-driven Human Mobility Modeling: A Survey and Engineering Guidance for Mobile Networking. *ACM Computing Surveys*, *48*(3), 38. doi:10.1145/2840722

Hosseini, S., & Li, L. T. (2016). Point-Of-Interest Recommendation Using Temporal Orientations of Users and Locations. *Lecture Notes in Computer Science*, *9642*, 330–347. doi:10.1007/978-3-319-32025-0_21

Huang, L., Ma, Y., & Liu, Y. (2015). Point-of-Interest Recommendation in Location-Based Social Networks with Personalized Geo-Social Influence. *China Communications*, *12*(12), 21–31. doi:10.1109/CC.2015.7385525

Huang, Q., & Wong, D. W. (2015). Modeling and visualizing regular human mobility patterns with uncertainty: An example using Twitter data. *Annals of the Association of American Geographers*, *105*(6), 1179–1197. doi:10.1080/00045608.2015.1081120

Jaccard, P. (1901). Distribution de la flore alpine dans le bassin des Dranses et dans quelques régions voisines. *Bulletin de la Société Vaudoise des Sciences Naturelles*, *37*, 241–272.

Jackson, K. T. (1996). All the World's a Mall: Reflections on the Social and Economic Consequences of the American Shopping Center. *The American Historical Review*, *101*(4), 1111–1121. doi:10.2307/2169636

Johnson, S. C. (1967). Hierarchical clustering schemes. *Psychometrika*, *32*(3), 241–254. doi:10.1007/BF02289588 PMID:5234703

Jurdak, R., Zhao, K., Liu, J., AbouJaoude, M., Cameron, M., & Newth, D. (2015). Understanding Human Mobility from Twitter. *PLoS One*, *10*(7), e0131469. doi:10.1371/journal.pone.0131469 PMID:26154597

Ketelaar, P. E., Bernritter, S. F., van't Riet, J., Hühn, A. E., van Woudenberg, T. J., Müller, B., & Janssen, L. (2017). Disentangling location-based advertising: The effects of location congruency and medium type on consumers' ad attention and brand choice. *International Journal of Advertising*, *36*(2), 356–367. doi:10.1080/02650487.2015.1093810

Kirchner, T. R., Cantrell, J., Anesetti-Rothermel, A., Pearson, J., Cha, S., & Kreslake, J. et al.. (2012). Individual mobility patterns and real-time geo-spatial exposure to point-of-sale tobacco marketing. In *Proceedings of the Conference on Wireless Health 2012* (p. 8). doi:10.1145/2448096.2448104

Liang, X., Zhao, J., & Xu, K. (2015). A general law of human mobility. *Science China. Information Sciences*, *58*(10), 1–14. doi:10.1007/s11432-015-5402-y

Lichtenstein, A. H., Appel, L. J., Brands, M., Carnethon, M., Daniels, S., & Franch, H. A. et al.. (2006). Diet and Lifestyle Recommendations Revision 2006: A Scientific Statement From the American Heart Association Nutrition Committee. *Circulation*, *114*(1), 82–96. doi:10.1161/CIRCULATIONAHA.106.176158 PMID:16785338

Lin, D. (1998). An information-theoretic definition of similarity. In *Proceedings of the 15th International Conference on Machine Learning* (pp. 296–304).

Liu, B., Xiong, H., Papadimitriou, S., Fu, Y., & Yao, Z. (2015). A general geographical probabilistic factor model for point of interest recommendation. *IEEE Transactions on Knowledge and Data Engineering*, *27*(5), 1167–1179. doi:10.1109/TKDE.2014.2362525

Liu, Y., Sui, Z., Kang, C., & Gao, Y. (2014). Uncovering Patterns of Inter-Urban Trip and Spatial Interaction from Social Media Check-In Data. *PLoS One*, *9*(1), e86026. doi:10.1371/journal.pone.0086026 PMID:24465849

Majid, A., Chen, L., Chen, G., Mirza, H. T., Hussain, I., & Woodward, J. (2013). A context-aware personalized travel recommendation system based on geotagged social media data mining. *International Journal of Geographical Information Science*, *27*(4), 662–684. doi:10.1080/13658816.2012.696649

Mok, D., Wellman, B., & Carrasco, J. (2010). Does Distance Matter in the Age of the Internet? *Urban Studies (Edinburgh, Scotland)*, *47*(13), 2747–2783. doi:10.1177/0042098010377363

Mozaffarian, D., Hao, T., Rimm, E. B., Willett, W. C., & Hu, F. B. (2011). Changes in Diet and Lifestyle and Long-Term Weight Gain in Women and Men. *The New England Journal of Medicine*, *364*(25), 2392–2404. doi:10.1056/NEJMoa1014296 PMID:21696306

Nekola, J. C., & White, P. S. (1999). The distance decay of similarity in biogeography and ecology. *Journal of Biogeography*, *26*(4), 867–878. doi:10.1046/j.1365-2699.1999.00305.x

Noulas, A., Scellato, S., Lambiotte, R., Pontil, M., & Mascolo, C. (2012). A tale of many cities: Universal patterns in human urban mobility. *PLoS One*, *7*(5), e37027. doi:10.1371/journal.pone.0037027 PMID:22666339

Palan, K. M., Gentina, E., & Muratore, I. (2010). Adolescent consumption autonomy: A cross-cultural examination. *Journal of Business Research*, *63*(12), 1342–1348. doi:10.1016/j.jbusres.2010.01.001

Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, *22*(10), 1345–1359. doi:10.1109/TKDE.2009.191

Pappalardo, L., Simini, F., Rinzivillo, S., Pedreschi, D., Giannotti, F., & Barabási, A.-L. (2015). Returners and explorers dichotomy in human mobility. *Nature Communications*, *6*(1), 8166. doi:10.1038/ncomms9166 PMID:26349016

Saad, U., Afzal, U., El-Issawi, A., & Eid, M. (2017). A model to measure QoE for virtual personal assistant. *Multimedia Tools and Applications*, *76*(10), 12517–12537. doi:10.1007/s11042-016-3650-5

Schein, A. I., Popescul, A., Ungar, L. H., & Pennock, D. M. (2002). Methods and Metrics for Cold-Start Recommendations. In *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 253–260). doi:10.1145/564376.564421

Schneider, C. M., Belik, V., Couronné, T., Smoreda, Z., & González, M. C. (2013). Unravelling daily human mobility motifs. *Journal of the Royal Society, Interface*, *10*(84), 20130246. doi:10.1098/rsif.2013.0246 PMID:23658117

Song, C., Qu, Z., Blumm, N., & Barabási, A.-L. (2010). Limits of Predictability in Human Mobility. *Science*, *327*(5968), 1018–1021. doi:10.1126/science.1177170 PMID:20167789

Subrahmanian, V. S., & Kumar, S. (2017). Predicting human behavior: The next frontiers. *Science*, *355*(6324), 489. doi:10.1126/science.aam7032 PMID:28154052

Tay, L., & Diener, E. (2011). Needs and subjective well-being around the world. *Journal of Personality and Social Psychology*, *101*(2), 354–365. doi:10.1037/a0023779 PMID:21688922

Wang, J., de Vries, A. P., & Reinders, M. (2006). Unifying user-based and item-based collaborative filtering approaches by similarity fusion. In *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 501–508). doi:10.1145/1148170.1148257

Wang, S. S., & Stefanone, M. A. (2013). Showing off? Human mobility and the interplay of traits, self-disclosure, and Facebook check-ins. *Social Science Computer Review*, *31*(4), 437–457. doi:10.1177/0894439313481424

Wei, Y., Zheng, Y., & Yang, Q. (2016). Transfer Knowledge between Cities. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1905–1914). doi:10.1145/2939672.2939830

Wu, L., Zhi, Y., Sui, Z., & Liu, Y. (2014). Intra-Urban Human Mobility and Activity Transition: Evidence from Social Media Check-In Data. *PLoS One*, *9*(5), e97010. doi:10.1371/journal.pone.0097010 PMID:24824892

Yan, X.-Y., Han, X.-P., Wang, B.-H., & Zhou, T. (2013). Diversity of individual mobility patterns and emergence of aggregated scaling laws. *Scientific Reports*, *3*(1), 2678. doi:10.1038/srep02678 PMID:24045416

Yang, D., Zhang, D., Chen, L., & Qu, B. (2015). NationTelescope: Monitoring and visualizing large-scale collective behavior in LBSNs. *Journal of Network and Computer Applications*, *55*, 170–180. doi:10.1016/j.jnca.2015.05.010

Yang, D., Zhang, D., & Qu, B. (2016). Participatory Cultural Mapping Based on Collective Behavior Data in Location-Based Social Networks. *ACM Transactions on Intelligent Systems and Technology*, *7*(3), 30. doi:10.1145/2814575

Yang, Z., Lian, D., Yuan, N. J., Xie, X., Rui, Y., & Zhou, T. (2017). Indigenization of urban mobility. *Physica A*, *469*, 232–243. doi:10.1016/j.physa.2016.11.101

Ye, M., Yin, P., Lee, W.-C., & Lee, D.-L. (2011). Exploiting geographical influence for collaborative point-of-interest recommendation. In *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 325–334). doi:10.1145/2009916.2009962

Yin, H., Cui, B., Zhou, X., Wang, W., Huang, Z., & Sadiq, S. W. (2016). Joint modeling of user check-in behaviors for real-time point-of-interest recommendation. *ACM Transactions on Information Systems*, *35*(2), 11. doi:10.1145/2873055

Yuan, J., Zheng, Y., & Xie, X. (2012). Discovering regions of different functions in a city using human mobility and POIs. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 186–194). doi:10.1145/2339530.2339561

Yuan, Q., Cong, G., Ma, Z., Sun, A., & Thalmann, N. M. (2013). Time-aware point-of-interest recommendation. In *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 363–372). doi:10.1145/2484028.2484030

Zhang, C., & Wang, K. (2016). POI recommendation through cross-region collaborative filtering. *Knowledge and Information Systems*, *46*(2), 369–387. doi:10.1007/s10115-015-0825-8

Zhang, J.-D., & Chow, C.-Y. (2015a). CoRe: Exploiting the personalized influence of two-dimensional geographic coordinates for location recommendations. *Information Sciences*, *293*, 163–181. doi:10.1016/j.ins.2014.09.014

Zhang, J.-D., Chow, C.-Y., & Li, Y. (2014). LORE: exploiting sequential influence for location recommendations. In *Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* (pp. 103–112). doi:10.1145/2666310.2666400

Zhang, J.-D., Chow, C.-Y., & Li, Y. (2015b). iGeoRec: A Personalized and Efficient Geographical Location Recommendation Framework. *IEEE Transactions on Services Computing*, *8*(5), 701–714. doi:10.1109/TSC.2014.2328341

Zhao, S., King, I., & Lyu, M. R. (2016b). A Survey of Point-of-interest Recommendation in Location-based Social Networks. https://arxiv.org/abs/1607.00647

Zhao, S., Zhao, T., Yang, H., Lyu, M. R., & King, I. (2016c). STELLAR: Spatial-Temporal Latent Ranking for Successive Point-of-Interest Recommendation. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence* (pp. 315–322).

Zhao, Y.-M., Zeng, A., Yan, X.-Y., Wang, W.-X., & Lai, Y.-C. (2016a). Unified underpinning of human mobility in the real world and cyberspace. *New Journal of Physics*, *18*(5), 053025. doi:10.1088/1367-2630/18/5/053025

Zhou, E., Huang, J., & Xu, X. (2015). A Point-of-Interest Recommendation Method Based on User Check-in Behaviors in Online Social Networks. In International Conference on Computational Social Networks, *LNCS* (Vol. *9197*, pp. 160–171). doi:10.1007/978-3-319-21786-4_14

ENDNOTES

[1] Gowalla was a location-based social network launched in 2007 and closed in 2012.

[2] As a geo-social networking and discovery site, Whrrl was launched in October 2007, and the Whrrl service discontinued in April 2011 for the takeover by Groupon.

Tao Xu is now a master student in the School of Computer Science, Wuhan University, China. He received his B.S. degree in computer science from the Hunan University of Science and Technology, China, and has been an international visiting student with the Department of Electronic and Computer Engineering, Lehigh University, USA. His research interests include machine learning and data mining. E-mail: WhuXuTao@outlook.com.

Yutao Ma (corresponding author) received his Ph.D. degree in computer science from Wuhan University, China. He is currently an Associate Professor in the School of Computer Science, Wuhan University. Dr. Ma was with the Institute of China Electronic System Engineering Corporation (Beijing) as a post-doctoral fellow and has been a visiting scholar in the Department of Electronic and Computer Engineering, Lehigh University, USA. His research focus is on the development of large-scale software service systems and complex networks. In addition to a member of the ACM and the IEEE, he is now a senior member of the CCF and a member of the CCF Technical Committee on Services Computing. He has also been a vice chair of the Services Society Young Scientists Forum (China) and a member of the CCF Youth Working Committee and the CCF Young Computer Scientists & Engineers Forum (Wuhan Branch). E-mail: ytma@whu.edu.cn.

Qian Wang is a full Professor with the School of Computer Science, Wuhan University, China. He received the B.S. degree from Wuhan University in 2003, the M.S. degree from Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences, China, in 2006, and the Ph.D. degree from Illinois Institute of Technology, USA, in 2012, all in electrical engineering. His research interests include wireless network security and privacy, cloud computing security, and applied cryptography. He is an expert under the "1000 Young Talents Program" of China. He is a co-recipient of the Best Paper Awards from IEEE ICNP 2011, WAIM 2014, and IEEE TrustCom 2016, as well as a co-recipient of the Best Student Paper Awards from IEEE ICDCS 2017. He is a Member of the IEEE and a Member of the ACM. E-mail: qianwang@whu.edu.cn.