Will You Come Back / Check-in Again? Understanding Characteristics Leading to Urban Revisitation and Re-check-in

ZHILONG CHEN, Beijing National Research Center for Information Science and Technology (BNRist), Department of Electronic Engineering, Tsinghua University, China
HANCHENG CAO, Department of Computer Science, Stanford University, United States
HUANGDONG WANG, Beijing National Research Center for Information Science and Technology (BNRist), Department of Electronic Engineering, Tsinghua University, China
FENGLI XU, Beijing National Research Center for Information Science and Technology (BNRist), Department of Electronic Engineering, Tsinghua University, China
VASSILIS KOSTAKOS, University of Melbourne, Australia
YONG LI*, Beijing National Research Center for Information Science and Technology (BNRist), Department of Electronic Engineering, Tsinghua University, China

Recent years have witnessed much work unraveling human mobility patterns through urban visitation and location check-in data. Traditionally, user visitation and check-in have been assumed as the same behavior, yet this fundamental assumption can be questionable and lacks supporting evidence. In this paper, we seek to understand the similarities and differences of visitation and check-in by presenting a large-scale systematic analysis under the specific setting of urban revisitation and re-check-in, which demonstrate people's periodic behaviors and regularities. Leveraging a localization dataset to model urban revisitation and a Foursqaure dataset to delineate re-check-in, we identify features concerning POI visitation patterns, POI background information, user visitation patterns, user preference and users' behavioral characteristics to understand their effects on urban revisitation and re-check-in. We examine the relationship between revisitation/re-check-in rate and the features we identify, highlighting the similarities and differences between urban revisitation and re-check-in. We demonstrate the prediction effectiveness of the identified characteristics utilizing machine learning models, with an overall ROC AUC of 0.92 for urban revisitation and 0.82 for re-check-in, respectively. This study has important research implications, including improved modeling of human mobility and better understanding of human behavior, and sheds light on designing novel ubiquitous computing applications.

CCS Concepts: • **Social and professional topics → Geographic characteristics**; • **Information systems →** *Data mining*;

Additional Key Words and Phrases: Revisitation, re-check-in, prediction, human mobility

*Corresponding author.

Authors' addresses: Zhilong Chen, Beijing National Research Center for Information Science and Technology (BNRist), Department of Electronic Engineering, Tsinghua University, Beijing, China; Hancheng Cao, Computer Science, Department of Computer Science, Stanford University, Stanford, United States; Huangdong Wang, Beijing National Research Center for Information Science and Technology (BNRist), Department of Electronic Engineering, Tsinghua University, Beijing, China; Fengli Xu, Beijing National Research Center for Information Science and Technology (BNRist), Department of Electronic Engineering, Tsinghua University, Beijing, China; Vassilis Kostakos, University of Melbourne, Australia; Yong Li, liyong07@tsinghua.edu.cn, Beijing National Research Center for Information Science and Technology (BNRist), Department of Electronic Engineering, Tsinghua University, Beijing, China.

2474-9567/2020/9-ART76 $15.00
https://doi.org/10.1145/3411812

ACM Reference Format:
Zhilong Chen, Hancheng Cao, Huangdong Wang, Fengli Xu, Vassilis Kostakos, and Yong Li. 2020. Will You Come Back / Check-in Again? Understanding Characteristics Leading to Urban Revisitation and Re-check-in. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 3, Article 76 (September 2020), 27 pages. https://doi.org/10.1145/3411812

1 INTRODUCTION

Understanding human mobility patterns has been a heated topic in both academia and industry. Previous work shows that human mobility demonstrates a high degree of spatial and temporal regularity, which can be generally modelled as a combination of periodic behavior and exploration [17]. While abundant literature has focused on discovering periodic patterns in human mobility [6, 8, 16, 26, 42], little is known about the fundamental difference between exploratory behavior and periodic behavior. In this paper, we address this lack in knowledge by asking the question: under what circumstances will users revisit the same place, *i.e.*, will a user at a given location exhibit periodic visitation behavior? Would it be different if the passive record changes to active sharing, *i.e.*, check-ins? Specifically, we seek to understand those characteristics pertinent to people's urban revisitation and re-check-in behavior, demonstrating their similarities and differences, and investigate the predictability of revisitation and re-check-in.

Understanding characteristics of urban revisitation and re-check-in is of significant importance, especially on a population scale. Firstly, it can lead to better knowledge of people's visitation and preferences, which contributes to better comprehension on rhythms of human mobility. Secondly, plenty of potential benefits can be brought to the application of ubiquitous systems, particularly in terms of recommendations. To be more specific, with the knowledge of whether a user is likely to revisit a Point of Interest (POI), better recommendations can be achieved by filtering out those POIs with past visits but with low chances of being revisited, as well as highlighting those with the highest likelihood [7]. Lastly, comparisons between urban revisitation and re-check-in can provide fascinating opportunities for better understanding of whether and how one can be used as a proxy for the other. For example, through analyzing discrepancies and similarities between urban revisitation and re-check-in, we can reach a more comprehensive view on how the selective and social nature of check-ins works. This can potentially lead to more knowledge on the fundamentals of human behavior.

In this paper, we focus on analyzing the predictability of people's urban revisitation behavior and re-check-in behavior. We here leverage a 1.5-month-long localization dataset to analyze the roles of different factors in determining whether or not people will revisit the same place, and a 2-year-long large-scale global Foursquare dataset to analyze how different factors can shape whether periodic patterns will show for check-ins. We here utilize POI-level features for locations so as to enhance the semantic meanings of visits. We analyze the relationship between revisitation/re-check-in and 1) POI-domain features, including the number of total records at the POI, number of people visiting/checking-in at the POI, the number of people revisiting/re-checking-in at the POI, POI category, and GDP, population and area of the country that the POI locates in; 2) user-domain features, including user's number of records/check-ins, activeness of visit and revisit, user's preference of POIs to check-in, visit and revisit, user's cross-category re-check-in rates, user's radius of gyration, and user's temporal visitation/check-in features.

For features in the POI domain, we discover that people's revisitation rate positively correlates with the total number of records at a POI, negatively correlates with the number of people visiting the POI and demonstrates a U-curve with the number of people revisiting the POI. People's re-check-in rate positively correlates with the total number of check-ins at a POI, the number of people checking in at the POI and the number of people re-checking in at the POI. We show that the re-check-in rate negatively correlates with the GDP of a country, while the scale of a country (population and area) does not show a significant correlation with the re-check-in rate. For features in the user domain, we identify positive relationships between people's check-in activeness and revisit activeness and re-check-in rate, while a U-shape is delineated between people's visit activeness and re-check-in rate. Strong preference for checking-in at and revisiting POIs pertaining to the categories of daily routines, for example, College & University and Outdoors & Recreation, shows a higher tendency of revisiting, while visit preference alone does not provide much information on re-check-in behaviors. The cross-category re-check-in rate can be indicative of re-check-in discovery in the way that people with a high re-check-in rate on POIs within a single category share higher re-check-in rate overall. What's more, we find out that higher mobility, more check-ins on weekends and at night can lead to a lower propensity of re-check-in. Through comparisons, we unravel the resemblances and discrepancies between revisitation and re-check-ins, highlighting the innate similarities and differences between passive and active location recordings which point to the nature of the two kinds of data sources.

Based on the insights derived from the identified characteristics, we build machine learning models to analyze the predictability of the identified characteristics. Taking all characteristics together, we reach an overall ROC AUC of 0.92 and 0.82 for revisitation and re-check-in respectively, which proves the high effectiveness of the features we extract and good predictability of space-related periodic behaviors.

The contribution of this paper can be summarized as follows:

- To the best of our knowledge, we are the first to address the problem of predicting, rather than modeling, urban revisitation behavior. We show that urban revisitation behavior is highly predictable through a number of explainable features derived from mobility records, which points to fundamental characteristics yet unknown in human mobility modeling.
- We introduce the idea of re-check-in and address their similarities and differences between revisitation, which gives rise to a better understanding of passive and active location records.
- We are the first to systematically analyze the roles of different factors in shaping urban revisitation and re-check-in. We exemplify how POI visitation records can indicate revisitation/re-check-in behaviors, and justify actions of periodicity negatively correlates with country GDP rather than country scale. We show how users' activeness, preference for POIs and behavioral traits can lead to revisitation/re-check-in discrepancies.
- We propose a prediction model of urban revisitation and re-check-in on the POI level. We verify the effectiveness of our model on a large-scale a Beijing localization dataset and a global check-in dataset, where we achieve accuracy of 0.92 and 0.82 ROC AUC in predicting POI revisitation behavior and re-check-in behavior, respectively. Our model can be leveraged in the future design of location-based ubiquitous computing systems.

2 RELATED WORK

Mobility Modeling. With regard to human mobility, studies have focused on estimating predictability based on stochastic models. Song et al. [30] derive the limit of predictability of mobility from the entropy of a user's historical trajectory using Fano's inequality [13, 14]. According to their calculation, the theoretical maximum prediction accuracy, referred to as potential predictability, achieves as high as 93% on a mobile phone collected dataset. The results indicate that human mobility actually implies huge regularity. Brockmann et al. [5] study the traveling behaviors of human based on the circulation of bank notes in the United States. They conclude that human movement can be accurately simulated by a two-parameter continuous-time random walk model. Furthermore, a number of works combine human mobility with social networks. By segregating similar users using the information from social media, more general and universal mobility patterns on a certain group of people can be extracted [12, 17]. On the other hand, Wesolowski et al. [33] suggest that mobility predictability is robust to the substantial biases in phone ownership across different geographical and socioeconomic groups. In summary, a large number of studies have discovered huge regularity in human movement patterns and concluded that human mobility is highly predictable, based on which numerous mobility models have been proposed [2, 17, 19, 22].

Mobility Prediction. Various statistical models have been proposed to predict people's future visitation. For example, Lu et al. [25] utilize the Markov Chain model to simulate users' transition patterns for visitation prediction. As a variant, hidden Markov model (HMM) is also widely used in human mobility modeling and visitation prediction. Mathew et al. [27] assume each latent state has a multinomial distribution over the locations. Various works [29, 41] improve HMM by integrating both user grouping and mobility modeling in the same model, in order to enhance the modeling ability and improve prediction accuracy. In addition, Dirichlet process is also widely adopted in predicting users' mobility. Jeong et al. [21] use Dirichlet process mixture model to cluster users based on their transition kernels. McInerney et al. [28] propose a hierarchical Dirichlet model, LocHDP, which shares temporal parameters within users and keeps spatial parameters unique to each user. More recently, researchers consider context-aware mobility modeling, *i.e.*, considering the information from external channels like social media content and social network graphs for better visitation prediction. For example, Zhang et al. [39] leverage the geo-tagged tweet to build a local event detection system, and Wang et al. [31] combine human mobility modeling with social network analysis to improve prediction accuracy. Different from them, our work focuses on the scenario of understanding revisiting, *i.e.*, whether an individual will come back to a certain place after first their encountering.

Periodic Pattern Mining and Revisitation Analysis. Periodic pattern mining is one important topic in time series analysis [18]. In terms of the periodic pattern of spatio-temporal data, periodic visitation is an important case of revisitation behavior. Giannotti et al. [16] propose T-patterns to address the problem of efficiently mining frequent spatio-temporal point sequences. Mamoulis et al. [26] study the problem of mining frequent periodic mobility patterns. Zheng et al. [42] aim to detect the frequent traveling path between fixed locations. Yuan et al. [38] propose a Bayesian non-parametric model to discover periodic mobility patterns by jointly modeling geographic and temporal information. Xu et al. [35] and Cao et al.[9] further propose methods to detect periodic temporal modes from mobility traces, which reflects user living pattern. Different from them, our work focuses on more general revisitation behavior including aperiodic revisitations, rather than only considering periodic patterns as in these works. Our work is closest to [6], where the authors first introduce the concept of revisitation into mobility analysis and show different urban revisitation patterns. However, this work falls short in revealing contributors/mechanisms of revisitation behavior. Complimentary to them, in this paper, we systematically analyze the role of multiple POI-domain and user-domain features derived from mobility records in shaping urban revisitation together with re-check-in, based on which we propose machine learning models to accurately predict urban revisitation behavior and re-check-in behavior on the level of an individual user. Furthermore, we explicitly show that features derived from periodicity greatly benefit mobility prediction/location recommendation, which points to the importance of taking the information into account in the design of future location-based systems.

3 PROBLEM OVERVIEW AND DATASET

3.1 Problem Overview

3.1.1 Urban Revisitation and Re-check-in Prediction. Consider the scenario that we visit a certain Point of Interest (POI). One interesting preliminary before studying the periodicity of our visits is to predict whether we will come back to the same place. Intuitively, numerous factors concerning people's behaviors and characteristics of POIs can impact whether an individual will revisit the same place. For example, if the POI is one's home or workplace, one will definitely return to the place. However, if the POI is a remote tourist attraction, there is a high likelihood that one does not have the intention of returning to the place at least for some time. Similar circumstances are for the user perspective, too. If an individual enjoys traveling and searching for new places for trials, their propensity of returning to a certain place is significantly low. However, if they are routine workers, there is a high likelihood that much regularity can be spotted from their mobility and a higher ration of revisitation can be expected.

Things would be much more complicated when considering re-check-in, *i.e.*, coming back and checking-in again at a POI. As pointed out by previous works such as [6, 11], check-ins can be biased and selective, and as a result, can differ from people's visits. For example, one may have a lower tendency of checking in and sharing his home due to privacy concerns [24]; for a single hotel, one may feel reluctant to check-in again and again at the same place for consecutive days to avoid being humdrum. Combining with people's aforementioned innate visitation patterns, a different pattern may appear.

3.1.2 Problem Definition. In this paper, we aim to model human mobility patterns from the perspective of urban revisitation and re-check-in at any given POI. To be specific, for urban revisitation, if we spot someone's presence at a certain place, we are interested in 1) whether they will return, and 2) what factors contribute to their probability of returning. For re-check-in, if we spot someone's check-in at a certain place, we are interested in 1) whether they will return and check-in again, and 2) what factors contribute to their probability of re-checking in.

Note that previous work has shown that most people's consecutive records at the same places happen within one year for check-ins and within one month for localizations [6]. Therefore, we define *revisitation* as individual's returning to the same place within a month's time on localization data, and *re-check-in* as an individual's returning to and checking in again at the same place within a year's time on check-in data. Our question at hand is to model 1) what factors can shape people's behaviors of revisitation and re-check-in: what kinds of POIs are more likely to be revisited and re-checked in, and what kinds of people are more likely to revisit and re-check in; 2) how revisitation and re-check-in are similar to and different from each other; and 3) how the proposed factors can be utilized to predict people's revisitation and re-check-in behaviors.

3.1.3 Problem Overview. To answer the aforementioned questions, we set out to analyze POI factors and user factors affecting revisitation and re-check-in. Our goal here is to uncover the relationship between actions of revisitation/re-check-in and different POI and user factors. Note that visits can be viewed as interactions between users and POIs. We here define an interaction between a user and a POI as a user-POI *pair*, and *revisitation rate* as the ratio that pairs with revisitation occupy among all pairs, *re-check-in rate* as the ratio that pairs with re-check-in occupy among all pairs. For POI factors, we analyze how revisitation/re-check-in rate correlates with 1) visitation patterns of POI: number of records at the POI, number of people visiting the POI and number of people revisiting the POI; 2) background information of POI: POI category and GDP, population and area of the country that POI locates in. For user factors, we analyze how revisitation/re-check-in rate correlates with 1) user's visitation activeness: user's number of records, number of POIs to visit and to revisit; 2) user's preference: user's check-in preference, visit preference, revisit preference and cross-category re-check-in rate; 3) user's behavioral characteristics: user's mobility (measured by their radius of gyration), and their temporal behaviors of daytime vs. night percentage and weekday vs. weekend percentage. A hierarchical demonstration of these factors is delineated in Fig. 1.

To further verify the strength of these identified features, we leverage these features to predict revisitation and re-check-in. The better a set of features can prediction actions of revisitation and re-check-in, the stronger that these features are in terms of forecasting revisitation and re-check-in.

3.2 Datasets

3.2.1 Localization Data in Beijing. This dataset was collected from one of the largest social networking and localization platforms in China. Whenever users execute third-party apps that utilize the platform's localization API, their WiFi connection, GPS and base station information is captured by the localization module, based on which their location is inferred and recorded. GPS information is utilized for localization when users are outdoors,

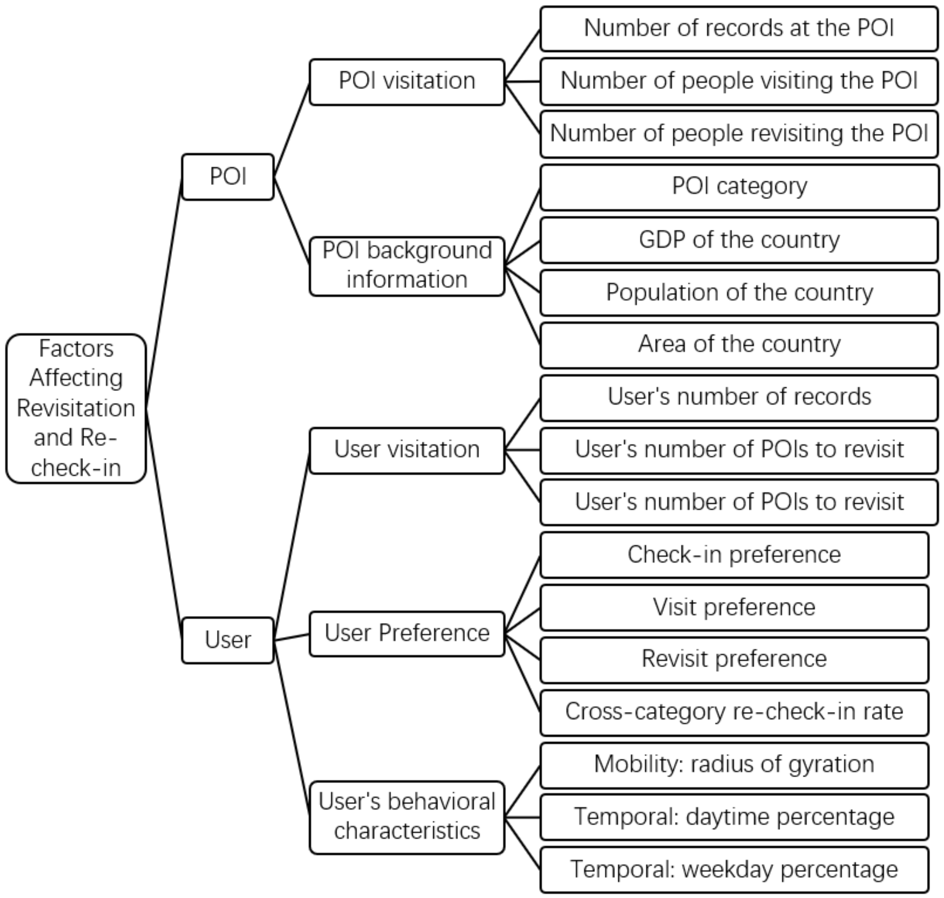


Fig. 1. Hierarchical demonstration of factors to analyze.

and when GPS is unavailable or users are indoors, more granular WiFi and base station information is used. It is reported that in more than 90% cases the location can be determined with an error of less than 22.5 meters [20], which ensures the accuracy of localization. Through matching users' actual locations to the spatial coverage of POIs, we obtain the actual POIs that users locate in. The dataset contains 3,097,863 localization records as a whole, incorporating 15,000 active users' localization records on 76,298 POIs between September 17th and October 31st, 2016 in Beijing. Each of these records contains the anonymized user ID, time, location, the associated POI, and POI type information. Here POIs are classified into 18 main categories including food, firm, entertainment, residence, school, shop, tourist attraction, transportation, etc.

3.2.2 Global-scale Check-in Data. This publicly available global-scale check-in dataset [36] consists of 22,809,624 check-in records on Foursquare, one of the largest location-based social networking services around the world. The dataset records 114,324 users' check-ins at 3,820,891 unique POIs between April 2012 and January 2014. Each record demonstrates one people's check-in at a certain place at a certain time, and contains an anonymized and unique user ID, venue ID of the POI, UTC time to check-in and timezone offset of the POI. Venue information and city information are additionally offered, which makes it possible to obtain GPS, category and country of venues and to match venues to cities.

Foursquare classifies POIs into 429 categories, and further clusters them into 10 major categories: Arts & Entertainment, College & University, Event, Food, Nightlife Spot, Outdoors & Recreation, Professional & Other Places, Residence, Shop & Service and Travel & Transport [1]. However, because there are no records belonging to the category of Event in our dataset, we exclude it from further consideration. We also collect GDP (Gross Domestic Product) information about countries, and population and area statistics about the countries in our analysis. GDP data is extracted from the International Monetary Fund public data [2]. Population and area of countries are gathered from open source dataset World Population Review [3].

3.2.3 Data Pre-processing. To improve the robustness of our analysis, we take several steps to remove noise and incomplete records. As mentioned by [6], there is the possibility that some users' several consecutive check-ins are actually generated through one visit to the same place. Therefore, we consider as separate check-ins only those consecutive check-ins by the same person at the same place that are more than 30 minutes apart.

Secondly, to minimize the possibility that not identifying POI-user pairs as an instance of revisitation is a result of not having a large enough dataset, we only consider those check-ins that first occurred in the first 12 months of our dataset and those localization records that first happened in the first 0.5 months. Recall that for urban revisitation, we subsequently investigate whether the same person will come back to the same place again within one month since their first visit at that POI; for re-check-in, we examine whether the same person will come back to and check in at the same place again within one year since their first visit at that POI. Therefore, we ensure that all 'first visits', *i.e.*, a person's first check-in/visit at a certain place, have the potential to be checked-in again within one year in Foursquare dataset or to be visited again within one month in localization dataset, and can be thus recorded as instances of re-check-in and revisitation.

Finally, to remove the potential bias associated with limited records, we filter out those POIs visited by fewer than 5 people and users with fewer than 5 records. Table 1 demonstrates details of the filtered data. Eventually, for localization data, 751,470 records by 11,573 users on 14,837 POIs are retained. A total of 64,626 user-POI pairs remain, among which 42,574 are with revisitation and 22,052 are without revisitation. The average revisitation rate of this dataset is 65.9%. For the Foursquare check-in dataset, 6,045,678 check-ins by 111,075 users on 301,078 POIs are retained for the Foursquare dataset. A total of 3,267,775 user-POI pairs remain, among which 926,900 are with re-check-in and 2,340,875 are without re-check-in. This dataset shares an average re-check-in rate of 28.4%. From the characteristics, we can discover discrepancies between re-check-in rate for re-check-in data and revisitation rate for localization data. This corroborates [6, 11]: because check-ins are more performative [11] and biased to record 'unusual' visits rather than regularities [6], check-ins are less likely to be generated periodically and the overall re-check-in rate is therefore expected to be lower compared to revisitation rate.

Table 1. Basic characteristics of filtered data.

Characteristic	Revisitation Data	Re-check-in Data
Total records remaining	751,470	6,045,678
Total POIs remaining	14,837	301,078
Total users remaining	11,573	111,075
Total pairs	64,626	3,267,775
Total pairs with revisitation/re-check-in	42,574	926,900
Total pairs without revisitation/re-check-in	22,052	2,340,875
Revisitation/Re-check-in rate	65.88%	28.36%

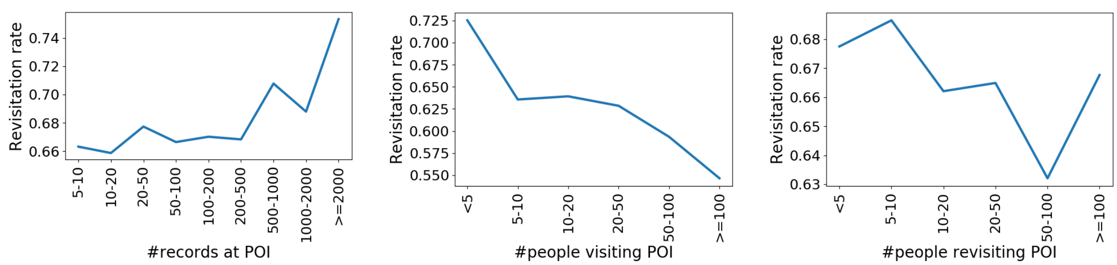
[1] https://developer.foursquare.com/docs/resources/categories
[2] https://www.imf.org/external/datamapper/NGDPD@WEO/OEMDC/ADVEC/WEOWORLD
[3] http://worldpopulationreview.com/

3.2.4 Ethical Considerations. Careful steps are taken to address the privacy issues concerning the sharing and mining of the data utilized in this paper. Firstly, the Foursquare dataset is publicly available, where users are anonymized. For the localization dataset, consent for research studies is included in the Terms of Use. User IDs are securely hash-coded so that anonymity is enhanced. Secondly, pre-processing and sanitization are carried out before our analysis of the data. Thirdly, our local university institutional board has reviewed and approved our research protocol. Finally, all data used in the study is stored in a secure off-line server. Access to the data is limited to only authorized members of the research team bound by strict non-disclosure agreements.

4 ANALYSIS: POI FACTORS AFFECTING REVISITATION AND RE-CHECK-IN

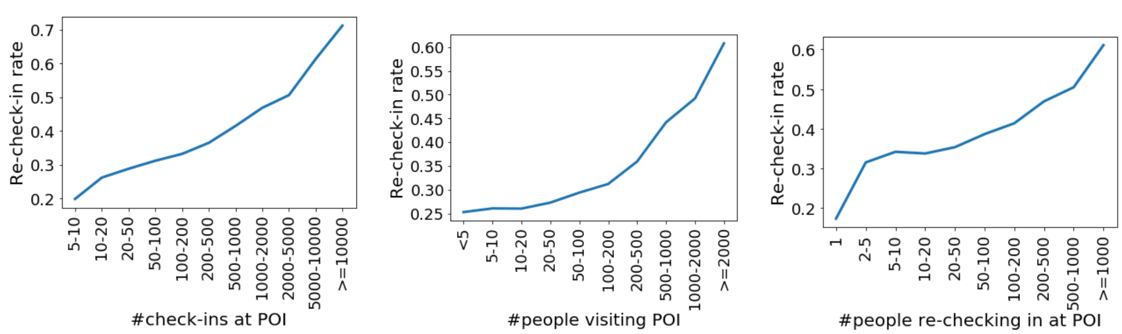
Many factors concerning characteristics both of POIs and users contribute to people's revisiting and re-checking in at a certain place. This section deals with the subset of features on the POI to be visited/checked-in. In particular, we investigate the influence of POI visitation patterns and POI background information on POI revisitation and re-check-in, respectively.

To uncover features affecting POI revisitation and re-check-in, we first consider POI visitation patterns. Specifically, we analyze the relationship between revisitation/re-check-in rate against: the total number of check-ins at the place, the total number of people who visit the place, and the total number of people to revisit the place.



(a) Number of records (b) Number of people to visit (c) Number of people to revisit

Fig. 2. The relationship between revisitation rate and number of records, visits, and revisits at POI on localization dataset.



(a) Number of check-ins (b) Number of people to visit (c) Number of people to revisit

Fig. 3. The relationship between re-check-in rate and number of check-ins, visits, and revisits at POI on check-in dataset.

Fig. 2(a) demonstrates the relationship between the average revisitation rate of POIs and the total number of records at the place. We observe that revisitation rate positively correlates with the number of records at the place on average. Specifically, POIs with 5 to 10 records share an average revisitation rate of 66.3%. This can increase to 75.4% for POIs with more than 2000 records. A similar tendency is discovered for re-check-in on the check-in dataset, where we discover a positive relationship between re-check-in rate and the number of check-ins at a POI. As shown in Fig. 3(a), re-check-in rate rises from 19.8% to 71.2% when the number of records at the POI increases from 5-10 to more than 10000. Intuitively, those POIs with more records can lead people to have the propensity to visit and possibly revisit the place. For example, a shopping mall within the center of the city may have a much higher likelihood to be revisited compared to a rural garden. However, some other popular POIs, for example, certain places of interest, do not tend to share a high likelihood of revisitation at least in the near future. This result suggests that POI popularity plays a dominant role in terms of revisitation rate. The trend is the same for re-check-in, too. In regard to the crowd's check-in numbers at a certain POI, we do expect that individuals tend to check in and re-check-in more at those places suitable for sharing, *i.e.*, places with more check-ins as a whole. In this way, a similar trend is to be expected.

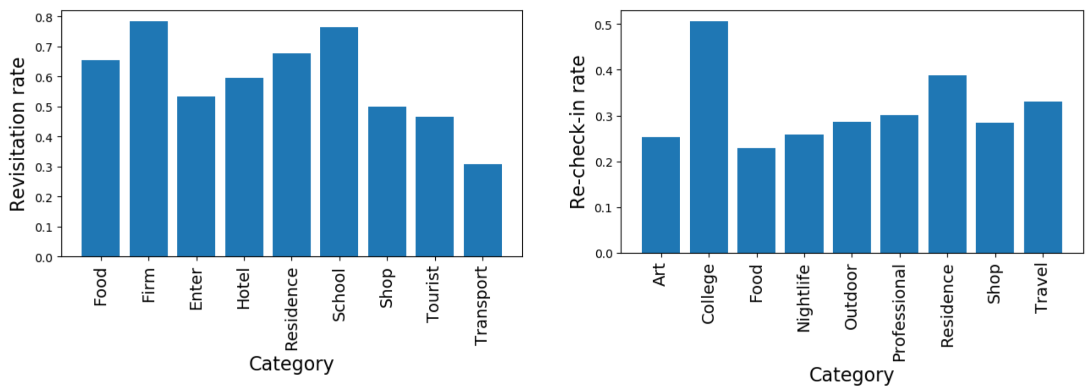
In terms of the number of people visiting a POI, as demonstrated in Fig. 2(b), the revisitation rate negatively correlates with the number of people visiting the POI. Specifically, the average revisitation rate of a POI drops from 72.5% to 56.2% when the number of people to visit there increases from fewer than 5 to more than 100. However, this is not the case for re-check-in. From Fig. 3(b), we observe a positive correlation between the number of unique people to check-in at a certain place and people's rate of re-check-in there. To be specific, the average re-check-in rate of a POI can rise from 26.1% to 60.8% when the number of people to check-in there increases from 2-4 to more than 2000. We believe that specially preferred places enjoy a higher level of stickiness, resulting in a higher revisitation rate. However, these results are likely due to unmotivated regularities, such as home/workplace, where users tend not to check in [11, 32], resulting in a lower rate of re-checking-in. As a result, discrepancies are shown.

Fig. 2(c) illustrates how revisitation rate changes with regard to the number of people to revisit a POI. A curve of U-shape can be spotted in terms of the relationship between revisitation rate and the number of people revisiting a POI for localization data, where the revisitation rate first drops when the number of people to revisit the POI declines from fewer than 5 to 50-100, but then increases when the number increases to 100 - 230. However, as shown in Fig. 3(c), a different trend of positive correlation can be identified for the relationship between re-check-in rate and revisit record at a POI. We discover that the more revisit records we observe at a certain place, the higher likelihood that more re-check-in behavior can be found per person, where the average re-check-in rate of POIs rises from 17.4% to 61.0% when revisit record increases from 1 to more than 1000. One way to explain the results is that both personal interests and public trends can influence revisitation/re-check-in. Those POIs particularly preferred are demonstrations of their special interests, while those POIs vastly revisited represent the preference of the public and have the chance of being worth revisiting. Balancing the two, a U-shape is to be expected. However, when the nature of check-ins limits those uninteresting or privacy-concerning POIs from being recorded and especially re-recorded, conformity of the public becomes more dominant. When observing others' re-check-in at a certain place, people have the tendency of following them, which can possibly lead to the same behavior of re-check-in.

4.1 POI Background

Background information about a POI can also have significant effects on revisitation and re-check-in. Here we analyze the impact of POI category and economic, cultural and scaling factors of the country where POI stands.

In Fig. 4 we show the average revisitation and re-check-in rate across different categories. When considering the relationship between revisitation rate and POI category for localization data, the revisitation rate across all categories are much higher than their corresponding results in terms of re-check-in rate. Speaking of relative heights, we discover that POIs related to daily routines, such as Firm, Residence and School, share a relatively

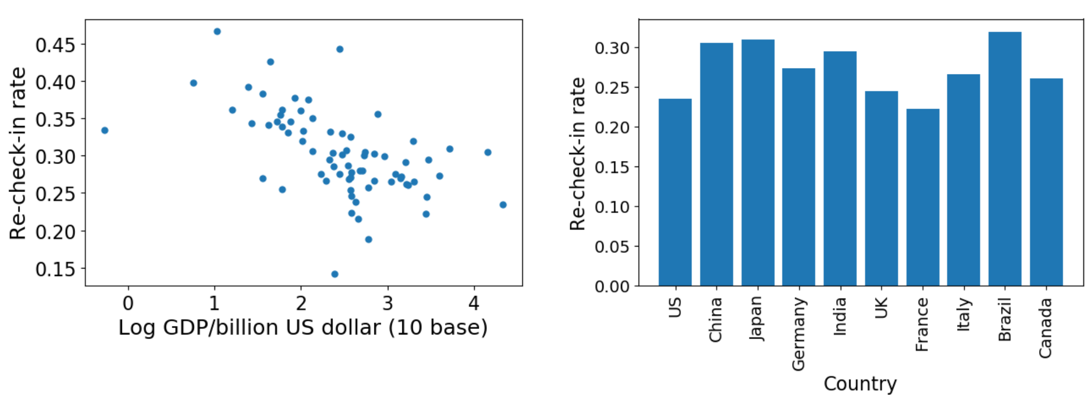


(a) Localization dataset

(b) Check-in dataset

Fig. 4. Revisitation and re-check-in rate across different categories for two datasets.

higher revisitation rate. The revisitation rates of POIs for Entertainment, Shop and Tourist are relatively low, and POIs pertaining to Transport share the lowest revisitation rate. In terms of re-check-in, we observe that POIs pertaining to daily routines share relatively higher re-check-in rate, too. The category of College & University has the highest re-check-in rate overall across all categories, and POIs of the type Residence have the second-highest average re-check-in rate. The results are easy to understand in the way that the vast majority of people living at school or College & University are students. Because these people's daily routines are relatively fixed, they have a higher likelihood of revisitation and re-check-in, resulting in high revisitation rates and re-check-in rates on average. Firm and Residence are places where people work and live. Provided that people transfer between workplaces and homes daily, a high likelihood of revisitation is to be expected. However, because of privacy and boredom concerns [11, 32], they can be removed from the choices of being checked in. Combining these aspects, we reach the result of a second-highest re-check-in rate.



(a) Re-check-in rate and GDP

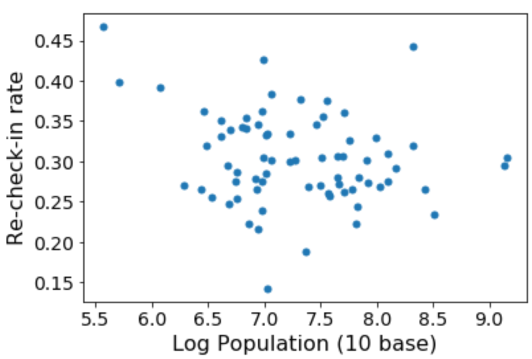
(b) Re-check-in rate of 10 countries with the largest GDP (in GDP descending order)

Fig. 5. Relationship between re-check-in rate and GDP of the country POI locates in.

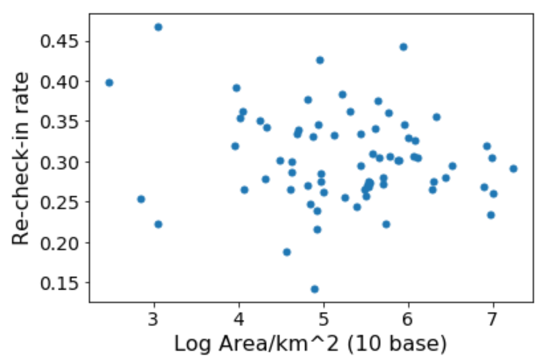
Cultural and economic factors like country can govern people's mobility patterns to a certain degree, and thus shaping people's revisitation and re-check-in behavior. Due to the limitation that it is hard to obtain large-scale worldwide localization data, we here only conduct experiments on the check-in dataset.

To identify the impact of these factors on re-check-in, we first analyze the relationship between average re-check-in rate and national Gross Domestic Product (GDP) of countries. As displayed in Fig. 5(a), average re-check-in rate of a nation negatively correlates with 10-base log GDP of the country with significance ($r = -0.547$, $p < 0.001$). This demonstrates that POIs in richer countries are less likely to be re-checked in overall. Several explanations can be provided to interpret the phenomenon. Firstly, people in richer countries can afford to travel and explore more. As noted by [3], people's income needs to cover the cost of moving in cities. The cost of movement positively correlates with distance, so people in richer countries generally can have a larger range of movement, leading to a relatively lower probability of re-check-in. Secondly, richer countries may have more places for recreation, for example, art museums and galleries. To be more precise, a positive correlation can be revealed between GDP with base 10 logarithm and the portion that the category of Arts and Entertainment takes ($r = 0.276$, $p = 0.019$). These places for entertainment are less likely to be revisited and thus re-checked in due to their nature of amusement as displayed in Fig. 4. As a result, POIs in richer countries have a lower likelihood to be re-checked in when more of these recreational places are taken into account, leading to the decrease of re-check-in rate. Thirdly, as demonstrated by [4], the number of infrastructures positively correlates with GDP. Therefore, the more prosperous a country is, the higher likelihood that more POIs appear and more POIs of the same kind appear. This can consequently result in increased competition between POIs, and from the perspective of users, they have a more diversified range of choices to satisfy the same need. As a result, they have less tendency of returning to the same POI but rather turn to other new POIs of similarity, which contributes to lower re-check-in rate.

To further consider the influence of countries' GDP, we examine the average re-check-in rate of the 10 countries with the largest GDP globally. As demonstrated in Fig. 5(b), it is not the case that re-check-in rates at countries with larger GDP have to strictly exceed those of countries with lower GDP. We attribute these differences to cultural factors. For example, it is likely that those who enjoy traveling are likely to have a lower re-check-in rate than people with highly structured routines.



(a) Re-check-in rate and population



(b) Re-check-in rate and area

Fig. 6. Relationship between re-check-in rate and population and area of the country POI locates in.

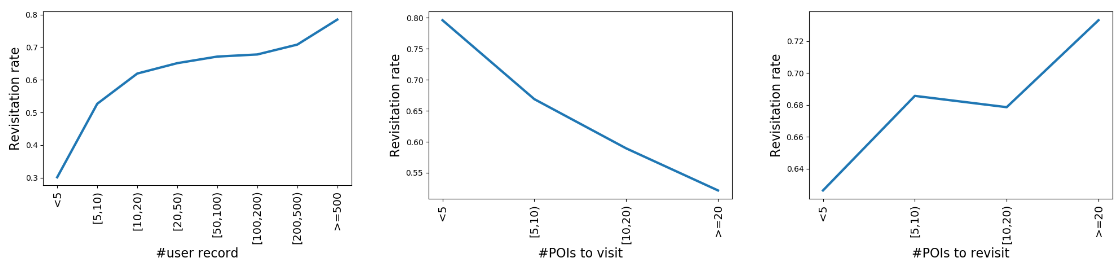
However, things change when speaking of countries' scale. Fig. 6(a) and Fig. 6(b) illustrate the relationship between re-check-in rate and countries' population and area, respectively. From the figures, not a significant correlation can be spotted either between re-check-in rate and population of countries ($r = -0.222$, $p > 0.05$)

or between re-check-in rate and area of countries ($r = -0.158$, $p > 0.05$). Therefore, we assert that the scale of countries does not prove to be a good indicator of re-check-in behaviors.

5 ANALYSIS: USER FACTORS AFFECTING REVISITATION AND RE-CHECK-IN

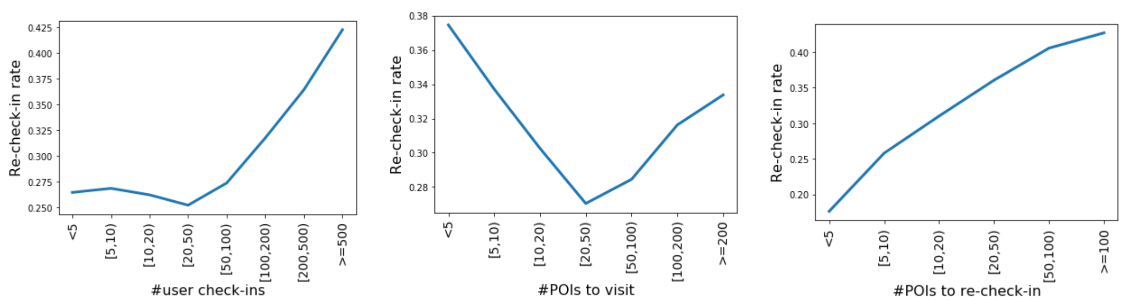
Next, we investigate how user characteristics can affect POI revisitation and re-check-in. Particularly, we study how users' visitation features, preference and behavioral characteristics can indicate POI revisitation and re-check-in.

Regarding users' own visitation features as a counterpart of POI visitation features, we hypothesize that they can also reveal revisitation and re-check-ins at any given POI. Specifically, we here analyze the relationship between users' POI revisitation and re-check-in and their total number of records/check-ins, total number of POIs visited and total number of POIs revisited.



(a) Number of records (b) Number of POIs to visit (c) Number of POIs to revisit

Fig. 7. The relationship between revisitation rate and number of records, number of POIs to visit and revisit on localization dataset.



(a) Number of check-ins (b) Number of POIs to check-in (c) Number of POIs to re check-in

Fig. 8. The relationship between re-check-in rate and number of check-ins, number of POIs to visit and revisit on check-in dataset.

Fig. 7(a) illustrates how revisitation rate shifts with regard to users' record number, where we discover a positive correlation between revisitation rate and users' number of records. Specifically, users with fewer than 5 records share an overall revisitation rate of 30.4%. As people's record number increases, the overall revisitation rate increases as well. For individuals with more than 500 records, their overall revisitation rate reaches 78.5%. A similar trend is demonstrated on the check-in data as well, which is depicted in Fig. 8(a). Users with fewer than 5 check-ins share an overall re-check-in rate of 26.5%. As people's check-in number increases, the overall re-check-in rate increases as well. For individuals with more than 500 check-ins, their overall re-check-in rate can increase to 42.3%. We believe that check-ins and records provide precious sources to model people's activeness and patterns of visitation. Provided that some kind of law underpins people's mobility patterns, we can identify the regularities of their lives from their check-ins and records. Consequently, we can observe more revisitation behavior from people with more records and more re-check-in behavior from people with more check-ins.

However, this is not the case for POI visitation. As exhibited in Fig. 7(b), we can discover a negative correlation between revisitation rate the number of POIs one visits. Specifically, people's propensity tp revisit the same POI is 79.6% when they visit fewer than 5 POIs. However, this number drops to 52.2% when their visits incorporate more than 20 POIs. For check-in data, a U-shape can be spotted when considering re-check-in rate against the number of POIs one visits (see Fig. 8(b)). To be more precise, people's propensity of re-checking in at the same POI drops from 37.5% to 27.0% when the number of POIs people visit increases from fewer than 5 to 20-50. However, the re-check-in rate then rises to 31.8% when the number grows to more than 100. Intuitively, more POIs to visit demonstrates more possible actions of explorations and less regularity. Therefore, a lower revisitation rate is to be expected. Speaking of the discrepancy between POI revisitation of localization data and re-check-in of check-in data, one possible explanation can be extracted from the bias of check-ins as mentioned by previous works [11, 32]. Users actively select whether to report a location and where to report. Considering the badge mechanism of Foursquare, it is reasonable that heavy users have higher tendency of chasing for badges, resulting in more re-check-in records. Combining this with general regularity, a U-shape is likely to be generated.

In terms of users' POI revisit number, a positive correlation can be observed between the overall POI revisitation rate and the number of POIs that individuals revisit (see Fig. 7(c)). For people revisiting fewer than 5 POIs, the possibility that they revisit a POI is 62.6%, while for people revisiting more than 20 POIs, their possibility of revisitation at a POI raises to 73.3%. A similar positive correlation can be identified from check-in data, where an increase from 17.6% to 40.8% on re-check-in rate can be discovered when the number of POIs users revisit increases from fewer than 5 to more than 50 (see Fig. 8(c)). We articulate that the number of POIs for a user to revisit/re-check-in can be a sign of people's mobility regularity/behavioral regularity. The more regular their lives are, the higher likelihood that they revisit the same POIs and the higher propensity that those revisited POIs take up a larger share of all of the POIs they visit, which consequently leads to higher revisitation rate and re-check-in rate.

5.1 Place Preference

Different people have distinct preferences for places to visit. For example, students have a higher tendency to visit those POIs regarding studies, while workers are more likely to visit factories and places of profession compared with students. Provided that previous work has proved the effectiveness of incorporating preference into location searching [37], it is reasonable for us to consider the influence of users' place preference on location revisitation and re-check-in here. Specifically, we investigate the impact of users' check-in preference, visit preference, revisit preference and cross-category re-check-in rate on their overall revisitation/re-check-in rate.

Here we only present the results on the check-in dataset to prevent the analysis from being too long. To extract features that capture people's preference, we leverage the aforementioned 9 major categories on Foursquare to establish users' preference vector. Each dimension of the vector illustrates users' normalized frequency of checking-in/having records of visiting/revisiting POIs pertaining to a certain category. For cross-category re-check-in rates, we calculate the re-check-in rate across the 9 major categories to demonstrate their relative preference across POIs with different functions.

To better illustrate users' differences and similarities, we aggregate users into clusters. Here we utilize the k-means algorithm to cluster users into groups with distinct characteristics, and leverage information of cluster centroids to demonstrate the extracted features. To search for the optimal value for k, we run k-means algorithms with k ranging from 2 to 20, calculate the sum of error between samples and cluster centroids based on 'Euclidean distance' and choose the elbow points, where not a significant drop of error sum exists, as our candidates for k.

Fig. 9. Centroid check-in preference.

Check-in Preference. For check-in preference, 6 disparate preference clusters of POI visitation are identified, all of which are shown in Fig. 9. Here cluster 1 demonstrates people who enjoy checking-in at Travel & Transportation the most, while people in cluster 2 have a high preference for College & University. People in cluster 3 have relatively balanced check-ins across different categories, while people in cluster 4 favor places for food. Cluster 5 is characteristic of those people fond of checking-in at places for Outdoors & Recreation, and cluster 6 of those who are keen on places belonging to the category of Professionals & Other Places.

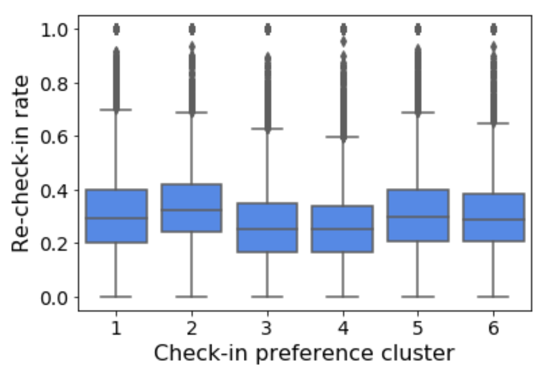


Fig. 10. Re-check-in rate across identified check-in preference clusters.

To uncover the relationship between people's check-in preference and re-check-in rate, we delineate the overall re-check-in rate across identified check-in preference clusters in Fig. 10. We discover that people in clusters 1, 2, 5 and 6 (*i.e.*, with higher tendencies of checking in at POIs concerning Travel & Transportation, College & University, Outdoors & Recreation and Professional & Other Places) have higher re-check-in rates. Conversely, people in clusters 3 and 4 (*i.e.*, with relatively balanced check-in preferences or with higher tendencies of choosing POIs for food as candidates for check-ins) have relatively lower re-check-in rates. We attribute this to the features of the POIs to be checked in. Provided that people have a preference for those categories which are more probable to be revisited, we are more likely to observe behaviors of periodicity on them, resulting in a higher re-check-in rate.

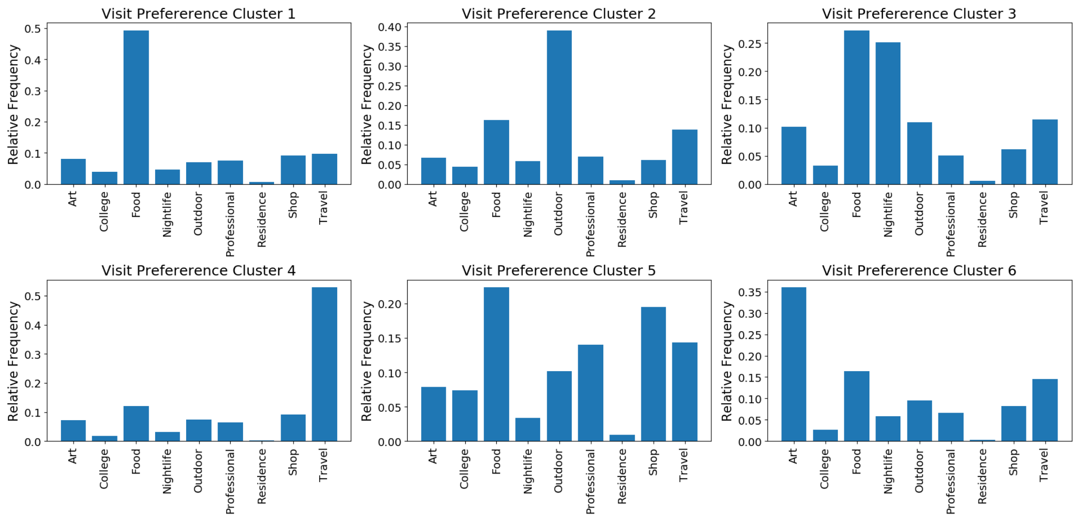


Fig. 11. Centroid visit/place preference.

Visit Preference. When it comes to people's choices of places to visit, our analysis identifies 6 distinct place preference types across people. People within cluster 1 show a strong preference for POIs relating to Food, and people belonging to cluster 2 are most likely to visit POIs for Outdoors & Recreation. A clear preference for Food and Nightlife Spot is exhibited in people in cluster 3, while the probability that people in cluster 4 visit POIs for Travel & Transport is particularly high. A relatively balanced visitation pattern lies within cluster 5, while cluster 6 share the feature that they are keen on visiting POIs belong to Arts & Entertainment.

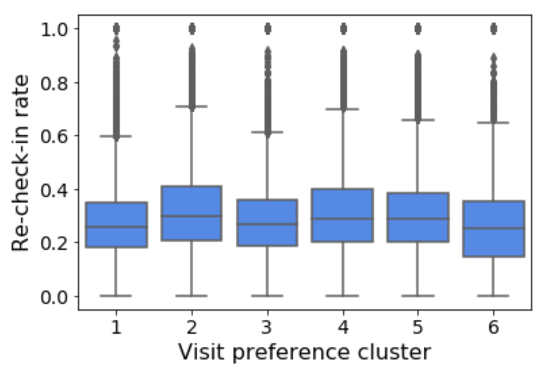


Fig. 12. Re-check-in rate across identified visit preference clusters.

Speaking of the relationship between people's preference of places to visit and re-check-in rate, the overall re-check-in rate of people in clusters 1, 3 and 6 are relatively lower than people in clusters 2, 4 and 5 (see Fig. 12). The overall re-check-in rate of clusters 1, 3 and 6 is 26.7%, 27.1% and 24.1%, respectively. However, the re-check-in rate across clusters 2, 4, and 5 is 30.7%, 30.3% and 29.3%. The results show that people's place preference of visitation correlates with re-check-in rates to a certain degree, but does not prove to be a reliable indicator of people's periodic behavior.

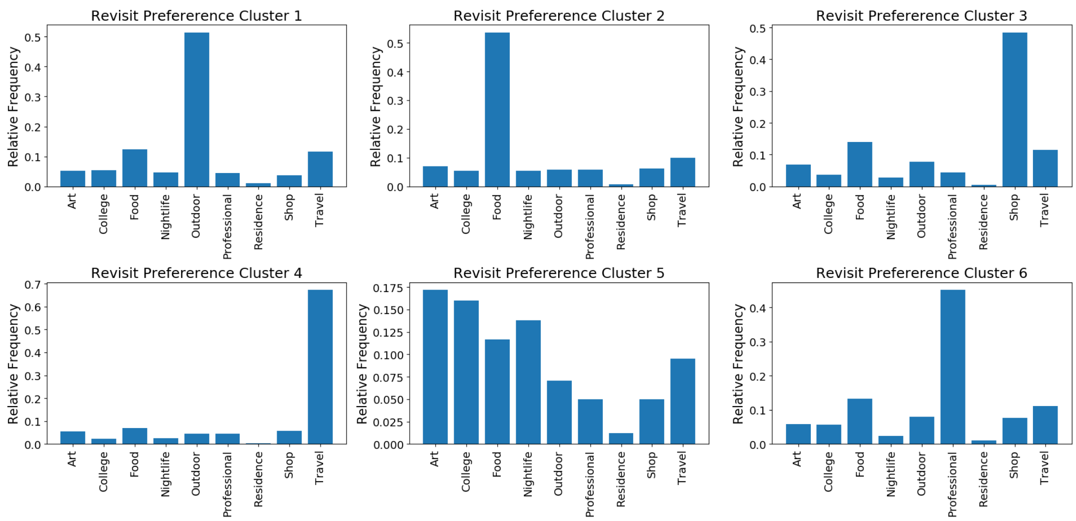


Fig. 13. Centroid revisit preference.

Revisit Preference. In terms of revisit preferences, as delineated in Fig. 13, 6 main clusters are discovered, each of which with a unique pattern. Cluster 1 is on behalf of the group of people who lay exceedingly interests in revisiting POIs for Outdoors & Recreation, while POIs for Food are the top priority of revisit for people in cluster 2. Preference of revisiting Shop & Service is a main characteristic of POIs in cluster 3, and that preference changes to the category of Travel & Transport for individuals in cluster 4. Cluster 5 represents those people with more balance revisit preferences, while cluster 6 is known as those people who have always been revisiting places in the realm of the category of Professional & Others Places.

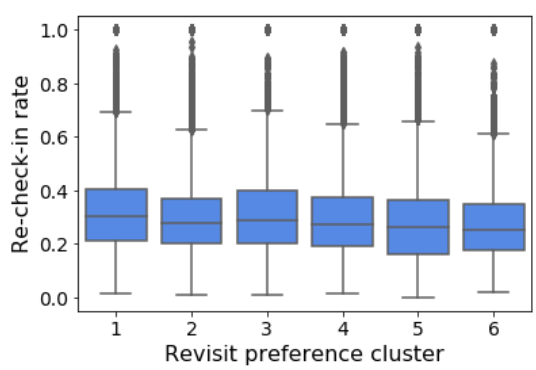


Fig. 14. Re-check-in rate across identified revisit preference clusters.

Fig. 14 illustrates the re-check-in rate across the 6 identified revisit preference clusters. We can observe that although the average re-check-in rates vary from cluster to cluster, no significant differences can be identified. Therefore, revisit preference does not prove to be a good informer of people's periodic behavior.

Cross-category Re-check-in Rate. Another demonstration of revisit/re-check-in preference is people's re-check-in rate across different kinds of POIs. As depicted in Fig. 15, we acquire once again 6 exemplary clusters

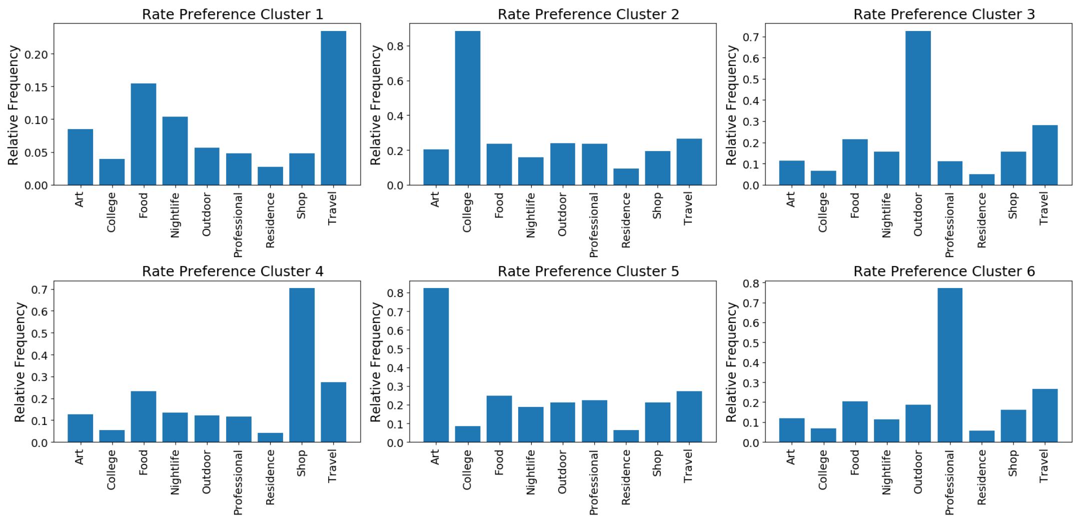


Fig. 15. Centroid re-check-in rate across categories.

of users, and huge discrepancies are shown across diverse clusters. To be specific, for people in cluster 1, the re-check-in rates for all categories are not so high, with POIs for Travel & Transport, Food and Nightlife Spot having relatively higher rates of re-check-in, while for individuals in cluster 2, POIs pertaining to College & University have the highest re-check-in rate. Re-check-in rate for individuals in cluster 3 is the highest for POIs concerning Outdoors & Recreation, and for individuals in cluster 4, the category with the highest re-check-in rate changes to Shop & Service. For people in cluster 5, the highest re-check-in rate can be spotted at places for Arts & Entertainment, while for individuals in cluster 6, they share the highest re-check-in rate for POIs in the scope of Professional & Other Places.

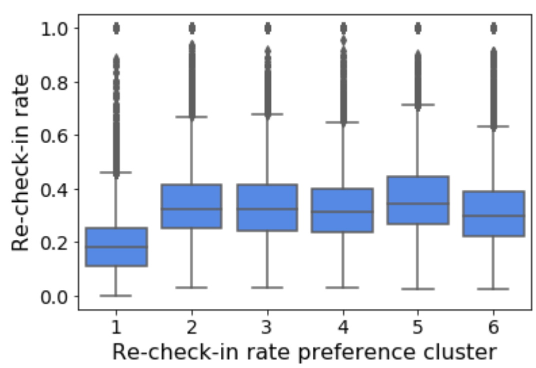


Fig. 16. Re-check-in rate across identified re-check-in rate preference clusters.

For the relationship between re-check-in rate and people's cross-category re-check-in rate preference, we plot the overall re-check-in rate of the 6 identified re-check-in rate preference clusters in Fig. 16. From the figure, we can see that people with high re-check-in rates at POIs concerning Arts and Entertainment share the highest re-check-in rate in general. A huge difference in the overall re-check-in rate can be spotted between cluster 1 and the other clusters. To be more precise, users belonging to cluster 1, with relatively high re-check-in rates POIs concerning Travel & Transport, Food and Nightlife Spot and relatively balanced and low re-check-in rate across all categories, have an overall re-check-in rate of 19.4%, which is only 0.68 times the average (28.4%). The result demonstrates that if we spot that people have a high re-check-in rate on certain categories, it is likely that their overall re-check-in rate is high as well. This gives rise to the possibility of portraying people's overall periodic behaviors from their periodic behaviors pertaining to certain categories of POIs.

5.2 Behavioral Characteristics

To further examine the influence of user's periodic behaviors, we turn to users' behavioral characteristics concerning mobility and temporality. Due to the limitation that certain fields are not provided in localization data, we are only able to analyze these aspects of check-in data. We here investigate the correlation between re-check-in rate and their radius of gyration, daytime check-in percentage, and weekday check-in percentage.

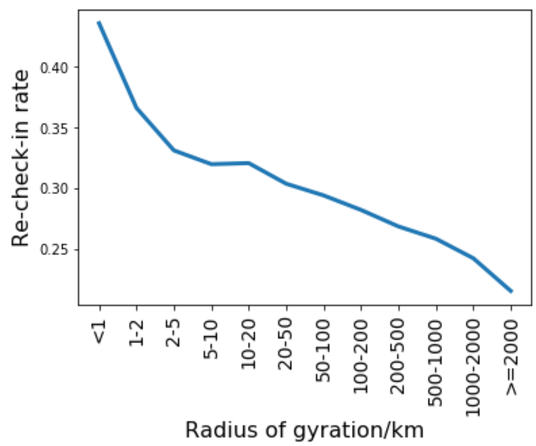
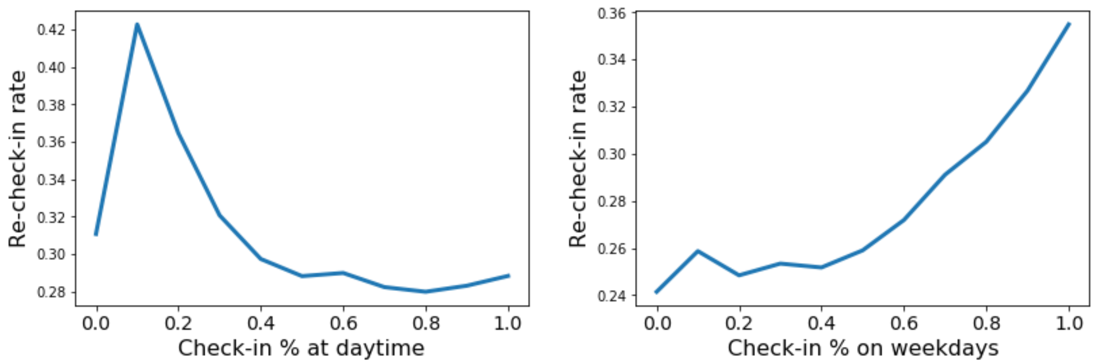


Fig. 17. Relationship between re-check-in rate and radius of gyration.

Radius of gyration, as suggested by [17], signifies as a user's characteristic distance of traveling. Fig. 17 demonstrates the relationship between re-check-in rate and user's radius of gyration. It can be concluded from the figure that people's radius of gyration negatively correlates with their re-check-in. For those people with a radius of gyration of less than 1 kilometer, their average re-check-in rate is 43.6%. As user's radius of gyration increases, their overall tendency of re-check-in drops significantly. For users with a more than 2000-kilometer radius of gyration, their re-check-in rate drops to 21.5%. We regard radius of gyration as a sign of people's mobility. People with larger radius of gyration tend to transfer more, enjoying lower levels of regularities in their trajectories. As a result, their inclination of returning to the same POI is lower due to their high mobility, resulting in lower re-check-in rates.

In terms of people's temporal behaviors, we consider their propensity of checking-in at daytime / at night and checking-in on weekdays / on weekends. Here daytime is defined as the interval between 8 a.m. to 8 p.m., and night is defined as 8 p.m. to 8 a.m. of the next morning. We delineate users' re-check-in rate with respect to their percentages of check-ins during daytime and on weekdays in Fig. 18(a) and Fig. 18(b). We find that people with relatively more check-ins at night share a relatively higher re-check-in rate. Specifically, people's re-check-in rate is the highest when approximately 10% of their check-ins are generated during day time. In terms of people's weekly visitation and check-in characteristics, people's overall re-check-in rate positively correlates with the ratio one checks-in on weekdays. People's re-check-in rate can increases from 24.1% to 35.5% when the ratio of their weekday check-ins rises from 0 to 1. Therefore, with a higher tendency of checking-in and transiting on weekdays, one has a relatively higher re-check-in rate compared to those who check-in and transit more



(a) Daytime percentage

(b) Weekday percentage

Fig. 18. Relationship between re-check-in rate and daytime & night and weekday & weekend check-in percentage.

on weekends. This corroborates the hypothesis that re-check-in records people's regularity. More check-ins on weekdays demonstrate more actions of daily routines. With a more regular daily routine, a higher re-check-in rate is therefore to be expected.

6 RESULTS: REVISITATION AND RE-CHECK-IN PREDICTION

On the basis of the aforementioned features to be analyzed, we utilize the identified characteristics to predict revisitation behaviors and re-check-in behavior with well-established machine learning techniques. Features concerning POIs as well as users are extracted on account of the characteristics mentioned in Section 4 and Section 5, and are binned into feature sets which are in accordance with the previous sections. Table 2 illustrates the features within each feature set.

Table 2. Description of feature sets.

Feature Set	Feature
POI	
POI visitation	#record, #people to visit, #people to revisit, average revisitation/re-check-in rate
POI background	category, GDP of country, population of country, area of country
User	
User visitation	#record, #POI to visit, #POI to revisit, average revisitation/re-check-in rate
User preference	record preference, visit preference, revisit preference, cross-category revitation/re-check-in rate
User's Behavioral characteristics	radius of gyration, record % at daytime, record % on weekdays

6.1 Experiment Setup

To verify the significance and influence of the identified features and feature sets, we utilize machine learning techniques that are as simple as possible and have good interpretability. We therefore report our results with

Logistic Regression Model, Naive Bayes Model, AdaBoost Model and Random Forest Model, respectively. In terms of our criterion for the justification of accuracy, we leverage the area under the receiver operating characteristic (ROC) curve (AUC) as an indicator of the probability that a randomly chosen positive instance with revisitation/re-check-in will be ranked higher than a negative instance that is randomly chosen. We here regard ROC AUC as a demonstration of the strength of the identified features. The higher our ROC AUC, the higher likelihood that revisitation/re-check-in information can be uncovered from the features and the more predicative the features are for revisitation/re-check-in prediction. Through 10-fold cross-validation, the ROC AUC for varied models with different features are estimated.

6.2 Predicting Revisitation and Re-check-in

Table 3 summarizes our results for revisitation prediction and re-check-in prediction.

Table 3. Revisitation prediction and re-check-in prediction performance ROC AUC through 10-fold cross-validation.

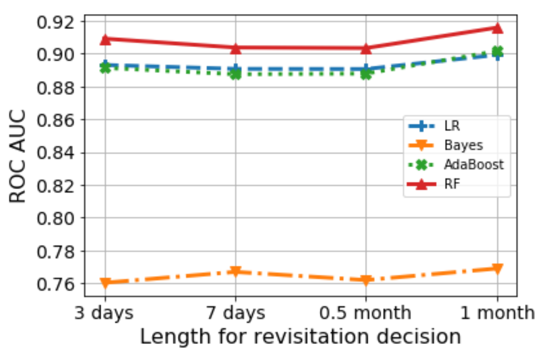
Model	Revisitation data				Re-check-in data			
	Logistic Regression	Naive Bayes	AdaBoost	Random Forest	Logistic Regression	Naive Bayes	AdaBoost	Random Forest
Random Baseline	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50
POI	0.81	0.72	0.81	0.80	0.77	0.73	0.76	0.67
POI visitation	0.80	0.79	0.80	0.79	0.77	0.76	0.76	0.67
POI background	0.59	0.59	0.59	0.59	0.57	0.56	0.57	0.57
User	0.81	0.73	0.81	0.77	0.67	0.63	0.66	0.65
User visitation	0.81	0.81	0.81	0.78	0.67	0.65	0.67	0.66
User preference	0.77	0.66	0.78	0.76	0.63	0.61	0.65	0.65
User's behavioral characteristics	-	-	-	-	0.55	0.54	0.55	0.56
All features	**0.90**	**0.77**	**0.90**	**0.92**	**0.80**	**0.71**	**0.80**	**0.82**

For revisitation prediction, we can see that ROC AUC for our identified features are similar across models, which shows the strength of our identified features. When we compare features in the POI domain and user domain, it can be figured out from the table that ROC AUC for POI (0.72-0.81) are almost the same as the user domain (0.73-0.81), which demonstrates that the identified POI features and user features share a same level of predictability. When considering specific feature sets, we discover that the subsets of features pertaining to POI visitation (0.79-0.80) and user visitation (0.78-0.81) perform the best across all identified feature sets for all four models. The model based on POI background performs the lowest ROC AUC of 0.59 for all four models, while for features regarding user preference, ROC AUC of 0.66-0.78 are shared. Therefore, we assert that POI visitation information and user visitation information are the most indicative of revisitation prediction.

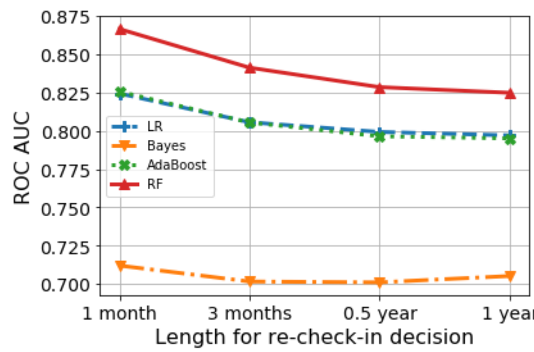
Combining all identified features, we can reach ROC AUC of 0.77-0.92 for the "full model". Specifically, we achieve the highest ROC AUC of 0.92 through Random Forest Model. A high ROC AUC of 0.90 can be obtained with simple models of the Logistic Regression Model and AdaBoost Model, which further corroborates the strengthen of the features we extract and the high predictability of urban revisitation. However, it is worth noting that the Random Forest Model does not prove to be the most indicative on all aspects. For example, for all user features, the highest ROC AUC is reached through the Logistic Regression Model and AdaBoost Model with ROC AUC of 0.81.

For re-check-in prediction, we discover similar results. Specifically, ROC AUC for our identified features are one again similar across models, which reiterates the strength of our identified features. When comparing features in the POI domain and user domain, we find that ROC AUC for POI (0.67-0.77) are higher than that in the user domain (0.63-0.66), which demonstrates that the identified POI features are better predictors than user features. When considering specific feature sets, we discover that the subset of features pertaining to POI visitation performs the best across all identified feature sets for all four models. For features concerning POI background, we reach a ROC AUC of 0.56-0.57 for four models. The model based on user visitation characteristics performs at 0.63-0.67 ROC AUC, while for features regarding user preference and user's behavior characteristics, relatively lower ROC AUC are identified, which are 0.61-0.65 and 0.54-0.56, respectively. Therefore, it can be indicated that POI visitation information is the most indicative of re-check-in prediction. What's more, the result that ROC AUC for user visitation is higher than that for user preference and user's behavioral characteristics illustrates that user visitation features are the most informative for re-check-in in the realm of characteristics concerning users.

Combining all features together, we reach ROC AUC of 0.71-0.82 for the "full model". Specifically, the highest ROC AUC can be reached with the Random Forest Model, with an overall ROC AUC of 0.82. Once again, the Random Forest Model is not the most indicative on all aspects. For example, for all POI features, the highest ROC AUC is reached through Logistic Regression Model with ROC AUC of 0.77.



(a) Localization dataset



(b) Check-in dataset

Fig. 19. ROC AUC of revisitation and re-check-in prediction with varying time for revisitation/re-check-in determination.

As a whole, all the models lead to satisfactory ROC AUC for revisitation/re-check-in predictions. Random Forest achieves the highest accuracy overall and is better at revisitation/recheck-in prediction modeling, while Naive Bayes leads to the lowest overall accuracy and is a relatively worse modeler. This can be interpreted in that of the four models, Random Forest Model is better at learning complex relationships that are highly non-linear and can more easily manipulate high-order data, while Naive Bayes is relatively simpler and can hardly learn the interactions between factors. Additionally, it is worth mentioning that users' behavioral characteristics are only available for re-check-in data and the inconsistency of feature number between two datasets may induce potential defects. However, we discover that removing users' behavioral characteristics from the analysis does not make much difference to the overall performance, which potentially indicates that the scenario would not substantially affect our comparison between datasets.

We further investigate the effect of the time period for re-check-in/revisitation on prediction performance. As mentioned in Section 3.1.2, the threshold for revisitation is defined as coming back within one month, and re-check-in as returning and checking-in within a year. By varying these thresholds, we investigate the ROC AUC when revisitation is defined as coming back within 3 days, 7 days, half a month, 1 month, and re-check-in as returning and checking-in again within 1 month, 3 months, half a year and one year. Fig. 19 shows how the

ROC AUC of revisitation and re-check-in prediction changes when varying these thresholds. We observe that both revisitation and re-check-in achieve similar prediction performance across different thresholds, suggesting that our analysis is robust. Random Forest achieves the best performance across all experiments, while Naive Bayes is the least successful. In addition, a slight drop is observed between the ROC AUC for check-in prediction and length for re-check-in decision. This indicates that long-term periodic behavior, especially long-term active re-check-ins, is more complicated and harder to model.

7 DISCUSSION

Several insights can be drawn from our prediction results. Firstly, the results demonstrate that both urban revisitation and re-check-in can be predicted with high accuracy, where better performances can be achieved through urban revisitation. This is easy to understand in the way that because check-ins are more performative [11], can be removed from further consideration to avoid boredom [24] and can be biased to record 'unusual' visits rather than regularities [6], relatively less regularity is demonstrated for check-in data compared to localization data. This can consequently raise the difficulty of prediction, resulting in relatively worse performances.

Second, the fact that combining POI features and user features can lead to better performance than each of the features alone shows the importance of considering both POI properties and user properties. This phenomenon is demonstrated in both revisitation modeling and re-check-in modeling. It is easy to understand that people's visiting and thus revisiting and re-checking in at a certain place can be regarded as an interaction between an individual and a POI. Therefore, identities of both ends of the interaction play an essential role in defining the satisfaction of the interaction and deciding whether later interactions, known as revisitation or re-check-in, can be expected.

Third, the results reveal that visitation characteristics can be more preditive than other features in terms of revisitation prediction and re-check-in prediction. To be more specific, the ROC AUC of POI visitation feature set alone is better than that of POI background feature set, and is approximately the same when these two feature sets are combined. Similarly, we identify certain cases from the user perspective, where user visitation features are more indicative than user preference and user's behavioral characteristics for revisitation prediction and re-check-in prediction. The performance of all user characteristics is almost the same when utilizing the sub feature set of user visitation. We ascribe this to the fact that these visitation features are within the same realm as revisitation and re-check-in. These features tend to be those concerning user-POI interactions. Therefore, it is likely that these features have a higher chance of being inter-correlated and more mutual information can possibly be obtained through these features, resulting in better performances for revisitation and re-check-in prediction.

Fourth, the place rather than who one may be, is more indicative of people's choice to repeat an act of checking-in to places. It is worth mentioning that features concerning POI visitation and user visitation are conceptually counterparts to each other. However, the results of our prediction demonstrate that for the Foursquare dataset of re-check-in, similar features from the perspective of POI and users can lead to substantial differences in the performance of our prediction; while for localization data of revisitation, not much difference can be identified. We can tell from the results that when modeling people's revisitation behavior, POI characteristics and user features show similar capability of inferring revisitation. However, when turning to check-ins, POI features prove to be better indicators of actions of re-check-ins. Combining the results, it is reasonable to attribute the discrepancies to check-ins' unique characteristics of active selections, *i.e.*, the place rather than the user himself is more indicative of their re-check-in choices.

Finally, the experiments show that the results are robust to changes in the thresholds for defining re-check-in and revisitation. This further demonstrates the strong regularities within people's periodic behaviors of revisitation and re-check-in. Additionally, the result suggests that predictability is somewhat improved for shorter-term revisitation/re-check-in predictions further justifies the high applicability of our discoveries: for cases where long-term records are hard to obtain, people's recent revisit/re-check-in behaviors can also be satisfactorily predicted.

7.1 Research Implications

Our study provides several insights for research in related areas. Firstly, our prediction experiment demonstrates that both revisitation and re-check-in can be predicted with high accuracy with the aforementioned features extracted. The obtained similarities demonstrate that because of the nature of check-in and localization data, they have the innate ability to model people's periodic behaviors. This indicates 1) the strength of the features we identify, 2) the innate similarity between revisitation and re-check-in and 3) the high predictability of people's periodic behaviors of revisitation and re-check-in.

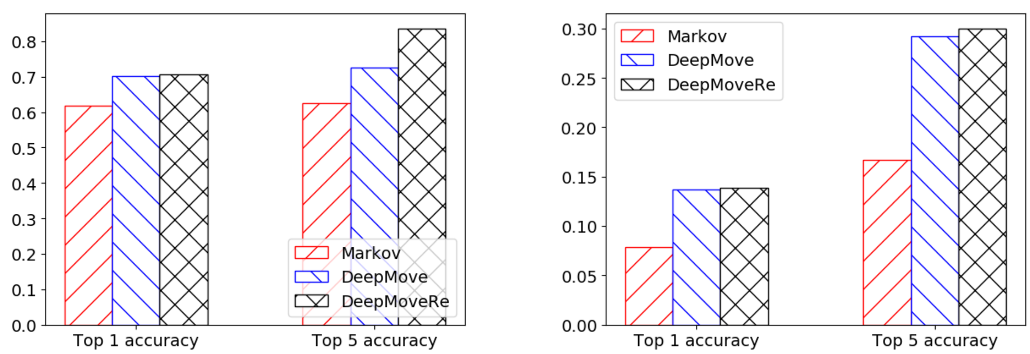
Meanwhile, it is important to call for attention on the differences between revisitation and re-check-in, which points to the usage of localization data and check-in data. Because localization data is often hard to obtain and limited in coverage, numerous works for research on human mobility have utilized check-in data, which are often open-sourced, accessible and globally covering, as proxies for localization data [10, 34, 40]. Our work reveals that certain discrepancies can occur in terms of results on localization data and check-in data because of their differences in nature. We articulate that more caution should be taken on directly generalizing the results on check-in data to localization data and represent localization with check-in data.

Secondly, our work contributes to the modeling of periodic patterns for POI visitation and check-in behaviors. To be specific, one of the preliminaries of periodicity is the existence of repetitive behaviors. Our work presents the first study on predicting whether records of revisitation will be observed, and demonstrates that accuracy of 0.92/0.82 for ROC AUC can be obtained with simple machine learning models based on the identified features for revisitation and re-check-in, respectively. Therefore, we articulate that it would be effective and efficient to predict the behavior of revisitation before modeling people's periodic patterns. Recall that the overall revisitation rate and re-check-in rate of all POI-user pairs is 28.4%/65.9%, and the action of ruling out POIs with approximately no possibilities for repetitive behaviors will substantially reduce the computational complexity for visitation periodicity modeling.

Thirdly, our work provides novel insight into the modeling of people's revisitation behavior. Prior literature has examined the action of revisitation in various formats, for example, mobile activity tracking app revisitation [23] and online crowdfunding site revisitation [1]. Our work on predictability on POI revisitation echos their work on people's intrinsic nature of re-connecting to the already familiar experiences. However, different from their work, we demonstrate that for POI revisitation, features of POIs rather than features of users can possibly be more indicative when considering corresponding features in the two domains, at least in our scenario. This contradicts those previous works on people's different kinds of revisitation behaviors to a certain degree. To be specific, our method of utilizing counterparts from POI perspective and user perspective enable better comparisons of the relative performance of characteristics from different angles. This potentially indicates that to better model interactions between POIs and users, extracting more characteristics from the POI perspective would probably be more efficient and effective.

7.2 Potential Applications

Valuable insights for application design can be drawn from our work. Firstly, our work benefits the applications concerning mobility prediction. Taking user's visitation history together with the location features, these applications can better model whether a user is likely to move to a certain place for a second time, which decides whether this place can be regarded as candidates of the next location for prediction. For example, we can add a revisitation/re-check-in filtering layer after the output layer of the state-of-the-art DeepMove mobility prediction model [15]. Taking 1) the prediction list with a descending ranking in prediction score, 2) people's historical trajectories, and 3) revisitation/re-check-in information together, the filtering layer decides whether those historically-encountered locations are likely to be revisited (for localization data) or re-checked-in (for check-in data). If not, the locations are removed from the top of the prediction list. In this way, we form a new prediction model named DeepMoveRe. Fig. 20 demonstrates its performance compared with existing models, where we delineate the top-1 prediction and top-5 prediction accuracy of baseline Markov model, DeepMove and DeepMoveRe on localization dataset and Foursquare check-in data, respectively. We observe that both top-1 and top-5 prediction accuracy can be improved when we take revisitation/re-check-in information into the consideration of the prediction models, which demonstrates the utility of revisitation and re-check-in prediction.



(a) Localization dataset

(b) Foursquare dataset

Fig. 20. Mobile prediction comparison between Markov model, DeepMove model and DeepMoveRe model.

Better prediction of human mobility and revisitation/re-check-ins can improve real-world scenarios in several ways. In smart transportation systems, improved predictions can give rise to better strategies for transport planning and management, such that traffic congestion can be eased. In the context of navigation systems, more appropriate routes can be identified when other people's mobility is better predicted and taken into consideration. Finally, with ride-sharing platforms, customer demand can be better understood and resources can be more efficiently allocated so that the overall demand of the crowd can be satisfied.

Secondly, our work gives rise to the better design of recommendation systems. Combining users' visitation histories and POI characteristics, an application can potentially better determine whether users have the inclination of returning to certain POIs. Therefore, when users come to a certain region, the application can perform better for places to recommend. For example, for a user who enjoys trying new places for food, we may extract his feature of being not willing to return to categories of food. Therefore, when the user comes back to a certain place, we can rule out those restaurants he has previously visited for our list of recommendations, which can consequently improve user experience.

Thirdly, our work can lead to improvements on applications for automatic event reminders. To be specific, leveraging users' visitation history and POIs' features of different kinds, applications of such kind can better determine whether a place is worthy of being added the to reminder list. If we successfully rule out the "redundant" candidates, a vast majority of noise can be removed and user experience can be significantly improved.

7.3 Limitations and Future Work

We still have a number of limitations and have to leave them for future work. Firstly, due to the limitation of data acquisition, our localization dataset only covers records of 15,000 users in 1.5 months. Therefore, long-term revisitation behaviors can hardly be managed. Although the Foursquare dataset has already crossed two years, because we wait 12 months for re-check-in prediction, the "valid" time of accumulating POI-user pairs is limited to 12 months, which takes up only a small portion of the dataset. For further analysis, it is worth extending the study to a longer period. Secondly, our localization dataset only consists of users in a single city of Beijing. Although the Foursquare dataset to be utilized is composed of check-ins across the globe, most of these check-ins are generated in western and well-developed cities. To better model human's all-round revisitation and re-check-in behaviors, more analysis incorporating developing countries should be conducted so as to achieve a more balanced and more thorough understanding. Third, due to the limitation of our datasets, demographic features such as gender and age are not taken into consideration of our prediction. Further studies can act on to optimize the extracted features to achieve better prediction of people's POI revisitation. Fourth, because the users and POIs of Foursquare and localization dataset are from different sources and the locations are scarcely overlapped geographically, we are not able to indicate how re-check-in can indicate revisitation (and vise versa). We leave it as future work to explore the interactions and transferability between revisitation and re-check-in. Our work has shown that although revisitation and re-check-in share similarities, they have differences that result from their nature. Therefore, we highlight that for further studies, researchers had better comprehensively consider these similarities and differences so as to better model human behaviors and should be cautious about directly using check-ins as proxies of people's mobility.

8 CONCLUSION

In this paper, we present the first analysis to date on the predictability of urban revisitation and re-check-in. We identify a number of factors from POI and user perspective and examine their correlation with revisitation and re-check-in. We analyze the similarities and differences between revisitation and re-check-in, reaching a better understanding of people's periodic behaviors. We leverage prediction model to examine the predictability of the identified characteristics and reach a high level of accuracy for prediction with our identified characteristics. Our work provides novel implications both for research and application, which can potentially benefit research for human mobility modeling and works concerning interactions between locations and people.

ACKNOWLEDGMENTS

This work was supported in part by The National Key Research and Development Program of China under grant 2018YFB1800804, the National Nature Science Foundation of China under U1936217, 61971267, 61972223, 61941117, 61861136003, Beijing Natural Science Foundation under L182038, Beijing National Research Center for Information Science and Technology under 20031887521, and research fund of Tsinghua University - Tencent Joint Laboratory for Internet Innovation Technology.

REFERENCES

[1] Tim Althoff and Jure Leskovec. 2015. Donor retention in online crowdfunding communities: A case study of donorschoose. org. In *Proceedings of the 24th international conference on world wide web*. International World Wide Web Conferences Steering Committee, 34–44.

[2] Fereshteh Asgari, Vincent Gauthier, and Monique Becker. 2013. A survey on human mobility and its applications. *arXiv preprint arXiv:1307.0814* (2013).

[3] Luís MA Bettencourt. 2013. The origins of scaling in cities. *science* 340, 6139 (2013), 1438–1441.

[4] Luís MA Bettencourt, José Lobo, Dirk Helbing, Christian Kühnert, and Geoffrey B West. 2007. Growth, innovation, scaling, and the pace of life in cities. *Proceedings of the national academy of sciences* 104, 17 (2007), 7301–7306.

[5] Dirk Brockmann, Lars Hufnagel, and Theo Geisel. 2006. The scaling laws of human travel. *Nature* 439, 7075 (2006), 462–465.

[6] Hancheng Cao, Zhilong Chen, Fengli Xu, Yong Li, and Vassilis Kostakos. 2018. Revisitation in Urban Space vs. Online: A Comparison across POIs, Websites, and Smartphone Apps. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 4 (2018), 156.

[7] Hancheng Cao, Jie Feng, Yong Li, and Vassilis Kostakos. 2018. Uniqueness in the city: Urban morphology and location privacy. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 2 (2018), 62.

[8] Hancheng Cao, Jagan Sankaranarayanan, Jie Feng, Yong Li, and Hanan Samet. 2019. Understanding Metropolitan Crowd Mobility via Mobile Cellular Accessing Data. *ACM Transactions on Spatial Algorithms and Systems (TSAS)* 5, 2 (2019), 8.

[9] Hancheng Cao, Fengli Xu, Jagan Sankaranarayanan, Yong Li, and Hanan Samet. 2019. Habit2vec: Trajectory Semantic Embedding for Living Pattern Recognition in Population. *IEEE Transactions on Mobile Computing* (2019).

[10] Eunjoon Cho, Seth A Myers, and Jure Leskovec. 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*. ACM, 1082–1090.

[11] Henriette Cramer, Mattias Rost, and Lars Erik Holmquist. 2011. Performing a check-in: emerging practices, norms and'conflicts' in location-sharing using foursquare. In *Proceedings of the 13th international conference on human computer interaction with mobile devices and services*. ACM, 57–66.

[12] Nathan Eagle, Aaron Clauset, and John A Quinn. 2009. Location Segmentation, Inference and Prediction for Anticipatory Computing.. In *Proceedings of the 23rd AAAI Conference on Artificial Intelligence (AAAI)*. 20–25.

[13] Agner Krarup Erlang. 1909. The theory of probabilities and telephone conversations. *Nyt Tidsskrift for Matematik B* 20, 6 (1909), 87–98.

[14] Robert M Fano and David Hawkins. 1961. *Transmission of information: A statistical theory of communications*. Vol. 29. 793–794 pages.

[15] Jie Feng, Yong Li, Chao Zhang, Funing Sun, Fanchao Meng, Ang Guo, and Depeng Jin. 2018. Deepmove: Predicting human mobility with attentional recurrent networks. In *Proceedings of the 2018 World Wide Web Conference*. International World Wide Web Conferences Steering Committee, 1459–1468.

[16] Fosca Giannotti, Mirco Nanni, Fabio Pinelli, and Dino Pedreschi. 2007. Trajectory pattern mining. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD)*. 330–339.

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

[18] Jiawei Han, Hong Cheng, Dong Xin, and Xifeng Yan. 2007. Frequent pattern mining: current status and future directions. *Data mining and knowledge discovery* 15, 1 (2007), 55–86.

[19] Samiul Hasan, Christian M Schneider, Satish V Ukkusuri, and Marta C González. 2013. Spatiotemporal patterns of urban human mobility. *Journal of Statistical Physics* 151, 1-2 (2013), 304–318.

[20] Tencent Incorporation. 2020. Introduction to Tencent location based services. Retrieved May. 15, 2020 from http://lbs.qq.com/guides/loc.html

[21] Jaeseong Jeong, Mathieu Leconte, and Alexandre Proutiere. 2016. Cluster-aided mobility predictions. In *Proc. INFOCOM*.

[22] Shan Jiang, Yingxiang Yang, Siddharth Gupta, Daniele Veneziano, Shounak Athavale, and Marta C González. 2016. The TimeGeo modeling framework for urban mobility without travel surveys. *Proceedings of the National Academy of Sciences* 113, 37 (2016), E5370–E5378.

[23] Zhiyuan Lin, Tim Althoff, and Jure Leskovec. 2018. I'll Be Back: On the Multiple Lives of Users of a Mobile Activity Tracking Application. In *Proceedings of the 2018 World Wide Web Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 1501–1511.

[24] Janne Lindqvist, Justin Cranshaw, Jason Wiese, Jason Hong, and John Zimmerman. 2011. I'm the mayor of my house: examining why people use foursquare-a social-driven location sharing application. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 2409–2418.

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

[26] Nikos Mamoulis, Huiping Cao, George Kollios, Marios Hadjieleftheriou, Yufei Tao, and David W Cheung. 2004. Mining, indexing, and querying historical spatiotemporal data. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining (KDD)*. 236–245.

[27] Wesley Mathew, Ruben Raposo, and Bruno Martins. 2012. Predicting future locations with hidden Markov models. In *Proc. UbiComp*.

[28] James McInerney, Jiangchuan Zheng, Alex Rogers, and Nicholas R Jennings. 2013. Modelling heterogeneous location habits in human populations for location prediction under data sparsity. In *Proc. UbiComp*.

[29] Hongzhi Shi, Yong Li, Hancheng Cao, Xiangxin Zhou, Chao Zhang, and Vassilis Kostakos. 2019. Semantics-aware hidden Markov model for human mobility. *IEEE Transactions on Knowledge and Data Engineering* (2019).

[30] Chaoming Song, Zehui Qu, Nicholas Blumm, and Albert-László Barabási. 2010. Limits of predictability in human mobility. *Science* 327, 5968 (2010), 1018–1021.

[31] Dashun Wang, Dino Pedreschi, Chaoming Song, Fosca Giannotti, and Albert Laszlo Barabasi. 2011. Human mobility, social ties, and link prediction. In *Proc. KDD*.

[32] Gang Wang, Sarita Y Schoenebeck, Haitao Zheng, and Ben Y Zhao. 2016. "Will Check-in for Badges": Understanding Bias and Misbehavior on Location-Based Social Networks. In *Tenth International AAAI Conference on Web and Social Media*.

[33] A Wesolowski, N Eagle, A. M. Noor, R. W. Snow, and C. O. Buckee. 2013. The impact of biases in mobile phone ownership on estimates of human mobility. *Journal of the Royal Society Interface* 10, 81 (2013), 20120986.

[34] Tong Xia and Yong Li. 2019. Revealing Urban Dynamics by Learning Online and Offline Behaviours Together. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 1 (2019), 30.

[35] Fengli Xu, Tong Xia, Hancheng Cao, Yong Li, Funing Sun, and Fanchao Meng. 2018. Detecting popular temporal modes in population-scale unlabelled trajectory data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 1 (2018), 46.

[36] Dingqi Yang, Bingqing Qu, Jie Yang, and Philippe Cudre-Mauroux. 2019. Revisiting user mobility and social relationships in lbsns: a hypergraph embedding approach. In *The World Wide Web Conference*. ACM, 2147–2157.

[37] Dingqi Yang, Daqing Zhang, Zhiyong Yu, and Zhiwen Yu. 2013. Fine-grained preference-aware location search leveraging crowdsourced digital footprints from LBSNs. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*. ACM, 479–488.

[38] Quan Yuan, Wei Zhang, Chao Zhang, Xinhe Geng, Gao Cong, and Jiawei Han. 2017. PRED: periodic region detection for mobility modeling of social media users. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining (WSDM)*. 263–272.

[39] Chao Zhang, Liyuan Liu, Dongming Lei, Quan Yuan, Honglei Zhuang, Tim Hanratty, and Jiawei Han. 2017. TrioVecEvent: Embedding-Based Online Local Event Detection in Geo-Tagged Tweet Streams. In *Proc. KDD*.

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

[41] Chao Zhang, Keyang Zhang, Quan Yuan, Luming Zhang, Tim Hanratty, and Jiawei Han. 2016. Gmove: Group-level mobility modeling using geo-tagged social media. In *Proc. KDD*, Vol. 2016.

[42] Yu Zheng, Lizhu Zhang, Xing Xie, and Wei-Ying Ma. 2009. Mining interesting locations and travel sequences from GPS trajectories. In *Proceedings of the 18th international conference on World wide web (WWW)*. 791–800.