Proceedings of the Fourteenth International AAAI Conference on Web and Social Media (ICWSM 2020)

Gravity of Location-Based Service: Analyzing the Effects for Mobility Pattern and Location Prediction

Keiichi Ochiai,[1,2] **Yusuke Fukazawa,**[1] **Wataru Yamada,**[1,2]
Hiroyuki Manabe,[1*] **Yutaka Matsuo**[2]
[1]NTT DOCOMO, INC., Japan
[2]The University of Tokyo, Japan
keiichi.ochiai@acm.org

Abstract

Predicting user location is one of the most important topics in data mining. Although human mobility is reasonably predictable for frequently visited places, novel location prediction is much more difficult. However, location-based services (LBSs) can influence users' choice of destination and can be exploited to more accurately predict user location even for new locations. In this study, we assessed the behavior difference for specific LBS users and non-users by using large-scale check-in data. We found a remarkable difference between specific LBS users and non-users (e.g., check-in locations) that had previously not been revealed. Then, we proposed a location prediction method exploiting the characteristics of check-in locations and analyzed how specific LBS usage influences location predictability. We assumed that users who use the same LBS tend to visit similar locations. The results showed that the novel location predictability of specific LBS users is up to 43.9% higher than that of non-users.

Introduction

With the rapid adoption of smartphones, location-based services (LBSs) are being widely used in daily life because location information is easy to accurately and automatically acquire through global positioning systems (GPS), Wi-Fi, etc. For example, Pokémon Go[1] released by Niantic in July 2016 is played worldwide. Given the popularity of LBSs, location prediction is one of the most active research topics in data mining. Previous research noted that human movement is highly predictable (Wang et al. 2015; Song et al. 2010; Lian et al. 2015; Cho, Myers, and Leskovec 2011; Li et al. 2010). In other words, people spend most of their time at home, in the workplace, and at a few frequently visited places, and they periodically transfer back and forth among these locations. However, location prediction for novel places remains difficult (Wang et al. 2015; Lian et al. 2015). Novel location prediction is an important topic, because users need information of novel locations more than that of regular places. In addition, it can be utilized to develop many exciting applications (e.g., personal assistants such as Google Now, and urban planning).

The content of most LBSs is associated with real-world locations. For example, in Pokémon Go, rare Pokémon types are more frequent in particular areas, and PokéStops, where a user can obtain items, are associated with real-world locations. Meanwhile, Instagram users tend to visit scenic spots to take stylish pictures. Thus, content distribution is believed to influence people's visiting behavior. In other words, the content of LBSs can make people to gravitate to specific locations. Therefore, there is a possibility that the performance of predicting new locations for specific LBS users can be improved by considering whether the target user uses a specific LBS or not and by mining app-specific places from the user's check-in history.

One possible solution for a novel location prediction is location recommendation because recommender systems calculate a prediction probability for each unvisited location (Lian, Zheng, and Xie 2013). Although various factors (such as geographical distance (Ye et al. 2011; Kurashima et al. 2013), temporal factors (Yuan et al. 2013; He et al. 2016), and social relationships (Ma, King, and Lyu 2009; Ye et al. 2011)) have been explored in existing studies of location recommendation for location-based social networks (LBSNs), the effect of mobile apps that a user uses has not been revealed.

In this study, we conduct an extensive quantitative analysis on mobility patterns of specific LBS users by using large-scale check-in data of LBSNs. We analyze the differences between particular LBS users and non-users regarding the number of check-ins, check-in locations, temporal dynamics, and distance between successive check-ins. Check-in data of Foursquare and Instagram from Twitter is used. On the basis of this analysis, we propose a novel location prediction method based on collaborative filtering (CF) that exploits the characteristics of check-in locations. Then, we analyze how specific LBS usage influences location predictability. We assumed that users who use the same LBS tend to visit similar locations.

The contributions of this study are the following:

- We conducted a quantitative experiment to investigate how LBSs influence mobility patterns by using large-

[*]Currently with Shibaura Institute of Technology

[1]http://pokemongo.nianticlabs.com/

scale check-in data. We selected Pokémon Go and Instagram as examples and found that specific LBS users and non-users have different mobility patterns.

- We proposed a novel location prediction method that exploits the characteristics of the tendency of check-in locations.
- We analyzed the effect of LBSs on location predictability through CF. The results showed that we can achieve 20.0% and 43.9% higher novel location predictability for users of Pokémon Go and Instagram, respectively, than non-users in terms of recall@5.

Related Work

Analysis of the Relationship between Mobility Behavior and Usage of Specific LBS

Existing studies can be categorized into quantitative and qualitative analyses. A quantitative analysis uses check-in data such as that of Foursquare or a wearable device to capture a user's mobility behavior. On the other hand, a qualitative analysis uses survey data based on questionnaires.

Quantitative Analysis of Mobility Behavior of Mobile Users Noulas *et al.* (Noulas et al. 2012) investigated human mobility patterns regarding distance using Foursquare check-in data and revealed that the point-of-interest (POI) density is an important factor in urban human mobility. Noulas *et al.* also investigated the temporal dynamics of a place network using Foursquare check-in data (Noulas et al. 2015). They revealed that place networks dynamically change over time and leveraged this finding for a link prediction task of a place network. Silva *et al.* (Silva et al. 2013) compared the distance between successive check-ins, the popularity of regions in cities, and temporal patterns between Foursquare users and Instagram users. They found that the popular regions are compatible, and the temporal patterns are similar during the same day but distinct for different days of the week. Tasse et al. (Tasse et al. 2017) reported that most people geotag in places they visit a few times.

Althoff *et al.* (Althoff, White, and Horvitz 2016) reported that Pokémon Go increases the physical activity of users. They used wearable sensors (Microsoft Band) and Bing's search queries to conduct their experiment. They inferred active Pokémon Go users from search queries and analyzed these users' activities. Based on estimation, active Pokémon Go users walked 1,473 additional steps per day, thereby increasing their average life expectancy by 41.4 days.

Qualitative Analysis of Mobility Behavior of Mobile Users Colley *et al.* (Colley et al. 2017) conducted a user survey of 375 Pokémon Go users and analyzed their geographical activity. They reported that almost 60% of Pokémon Go users had visited at least one new place when playing Pokémon Go. Another survey (Paavilainen et al. 2017) of 1,000 Finnish Pokémon Go users reported the same trend as regards visiting new locations.

However, no existing work has reported on the effect for location prediction.

Location Recommendation as Novel Location Prediction

In the recommender system community, novel location prediction is the same as location recommendation because a recommender system calculates the prediction probability for each unvisited location (Lian, Zheng, and Xie 2013; Wang et al. 2016). Thus, we use "recommendation" interchangeably with "prediction."

Location recommendation has been widely studied because it is an active research area (Ma, King, and Lyu 2009; Ye et al. 2011; Yuan et al. 2013; Kurashima et al. 2013; He et al. 2016). These studies can be categorized into three groups on the basis of the factors involved (Yu and Chen 2015): (1) Geographical, (2) Social, and (3) Temporal factors.

Geographical factors Ye *et al.* (Ye et al. 2011) used the idea that human movement follows a power law for POI prediction. They developed a combined user preference through user-based CF, social influence from friends, and geographical influence as mentioned above. Kurashima *et al.* (Kurashima et al. 2013) proposed a method called *Geo Topic Model* to simultaneously model a user's interest and activity area. They supposed that a user tends to visit locations geographically close to the locations he/she visited in the past. In addition, the visited location is influenced by the user's interests (e.g., a user who is interested in art is more likely to visit museums)

Social factors Ma *et al.* (Ma, King, and Lyu 2009) proposed friend-based CF. They assumed that users are easily influenced by their trusted friends. Hence a user balances his/her own preference against the recommendations from friends. Ye *et al.* (Ye et al. 2011) exploited the work of Ma *et al.* (Ma, King, and Lyu 2009) for POI prediction. Based on the hypothesis that friends are likely to go to similar locations, they weighted the check-in record on the basis of social connections and similarity of their check-in locations. Li *et al.* (Li et al. 2016) defined three types of friends: (1) Social Friends who have a connection on an SNS, (2) Location Friends who visit the same locations, and (3) Neighboring Friends who live in nearby homes. They integrated these three types of friends to predict POIs and incorporated latent check-in that is estimated because the user cannot visit all POIs.

Temporal factors Yuan *et al.* (Yuan et al. 2013) proposed a method to exploit temporal information for POI prediction because people tend to visit various places at different times of the day (e.g., people visit restaurants around noon). They computed the similarity between time slots to permit interpolation given that a user-POI matrix created for each time slot is very sparse. The POI recommendations of He *et al.* (He et al. 2016) utilize the idea that human behavior is periodic, and the visit frequency (days of the week) depends on the POI category. For example, universities are often visited on weekdays, whereas bars tend to be visited on Fridays and weekends. They predict the POIs by using tensor decomposition to extract such latent patterns.

Figure 1: Screenshot of Pokémon Go. (a) Capturing the Pokémon (left), (b) wandering the virtual world (center), and (c) PokéStops (right)

In these studies, user-based CF is adopted as the basic framework, and each influence (geographical, social, and temporal) is added to improve the prediction accuracy. To the best of our knowledge, no existing work has considered applying specific LBS usage logs for location prediction.

Because some recent recommendation studies have aimed at recommendations based on diversity, serendipity, etc. rather than just accuracy (Kaminskas and Bridge 2016), existing studies can be categorized by not only the incorporated factors but also considered metrics such as diversity, serendipity (Kunaver and Požrl 2017). Thus, we briefly review the related work. Zhang *et al.* proposed a POI specific recommendation method by considering POI availability and diversity (Zhang, Liang, and Wang 2016). The work of Boim *et al.* (Boim, Milo, and Novgorodov 2011) and Ho *et al.* (Ho, Chiang, and Hsu 2014) proposed diversity-aware recommendation methods which can be combined with any CF-based method. Because our proposed method is CF-based method, we can integrate these diversity-aware method with the proposed method.

Overview of Location-based Services

In our experiments, we selected Pokémon Go and Instagram as our two LBSs because they have the highest number of users. In addition, because Pokémon Go has digital content and Instagram has real-world content, we adopted these two services to investigate their content characteristics.

Pokémon Go

Pokémon Go is a location-based mobile game. Figure 1 shows a typical screenshot of the Pokémon Go game. Players are encouraged to wander around real world to capture Pokémon creatures bound to real-world locations (Figure 1 (b)). GPS is used for matching the player's real world location with the virtual world. To capture a Pokémon, the player needs to use items such as Poké Balls, and Berries (Figure 1 (a)) that can be obtained from PokéStops located at monuments, landmarks, etc. (Figure 1 (c)). The player can fight with other players at Gyms that are also located at real-world POIs. Therefore, physical movement is an important characteristic of this game.

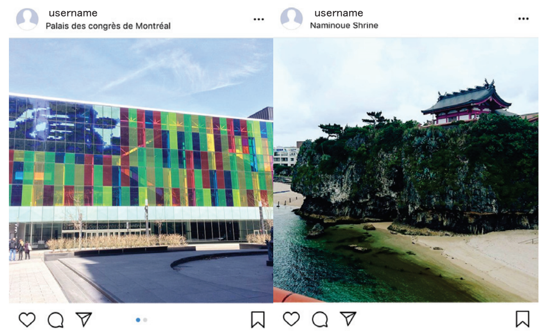


Figure 2: Screenshot of geo-tagged post of Instagram.

Instagram

Instagram is a photo-sharing social network service. A user can attach word-tags and geo-tags (i.e., location information) to the photos as shown in Figure 2. One motivation for the user to share photos is "Coolness," that is, to become popular or obtain more "Likes" (Sheldon and Bryant 2016). Because of these characteristics, users tend to share "Instagrammable" photos (i.e., photogenic scenes).

Definition and Dataset

Definition

In this research, we define the behavior of a user as the check-in history, which is a set of places he/she visited within a certain period of time. Let U be the set of all users, L be the set of locations, and T be the set of hourly-based time slots (i.e., a day is divided into 24 time slots). If each user is $u \in U$, each location is $l \in L$, and time is $t \in T$, then we define check-in and check-in history as follows.

Definition (Check-in) Check-in v_u is described by the user-location-time tuple (u, l, t), which indicates the user u visited location l at time t.

Definition (Check-in History) Assuming that the i-th check-in of the user u is $v_{u,i}$, then the check-in history of u is defined as $h_u = \{v_{u,1}, v_{u,2}, \cdots, v_{u,n}\}$

Dataset

Strategy behind the data collection We assume that the data must meet three requirements to validate our hypothesis: (1) the volume of check-in data is large enough to ensure statistical validity, (2) the data is publicly available for research reproducibility, and (3) the data has POI names for assured interpretation. Thus, Foursquare, Instagram, and Twitter are reasonable candidates. The terms of use of all these candidates, other than Twitter, prohibit data collection. Twitter has place-name tagging but the amount of data is small, whereas Foursquare and Instagram have POI names. Therefore, we collected check-in data of Foursquare and Instagram from tweets as they meet all the requirements. In addition, the location history data of Pokémon Go users is not publicly available because Niantic does not publish the API to access the data. Thus, we use service-related keyword matching to identify Pokémon Go users from tweets.

Details of dataset We use the check-in data of Foursquare and Instagram for almost half a year (from January 4, 2018, to July 15, 2018). The check-in data is collected from Foursquare and Instagram via Twitter. Figure 3 shows the flow of data collection and filtering. We obtain geotagged tweet data from Twitter API[2]. All geolocation is converted into a geographic grid such as geohash (Geohash 2008) and Grid Square (Japanese Statistics Bureau 1996). The reason we convert is that a POI integration is needed for analysis because POIs of Foursquare and Instagram are different. We used Grid Square (Japanese Statistics Bureau 1996) defined by the Statistics Bureau of Japan. The geolocation is represented by mesh code. We used the grid with 250-meter-long sides by considering a smartphone's GPS noise (von Watzdorf and Michahelles 2010). Subsequently, we extract tweets that are posted by Foursquare or Instagram on the basis of the source property because all tweets have a source property that indicates the client application of a tweeted post. Finally, we extract the POI from the tweets by using regular expressions because Foursquare tweets use the style of "I'm at POI NAME in CITY swarmapp.com/xxx" or "USER COMMENT (@ POI NAME in CITY) swarmapp.com/xxx", whereas Instagram tweets use the style of "USER COMMENT location: POI NAME instagram.com/xxx". The reasons we extract POI names from check-in are (1) to extract representative POIs from a geographic grid for interpretation, and (2) to filter out posts that have no fine-grained check-in locations because the location names in several posts are broader locations such as city or prefecture names. Specific app users are distinguished by the following rules:

- **Pokémon Go Users:** Users who meet at least one of the following conditions:
  (1) keywords such as "pokemon go" are in the user's profile text, or
  (2) a user who issued at least K tweets that include keywords such as "pokemon go," "pokestop," or several other service-related words (e.g., character names such as Pikachu and Eevee).
- **Instagram Users:** A user who has at least one tweet posted via Instagram. That is, the user first posted on Instagram, and then the post was automatically posted to Twitter.

We call the users who do not meet above conditions "others" or "non-users." Given the extraction of posts under these conditions, a check-in posted by a Pokémon Go user does not necessarily indicate a check-in posted while the user was playing Pokémon Go. Thus, not all check-in spots of Pokémon Go users are associated with Pokémon Go. On the other hand, because check-in data of Instagram users is collected by the source property, all check-in spots of Instagram users are associated with Instagram. Our purpose is to predict the locations to which specific LBS users go, not the locations where specific LBS users access a specific LBS.

Specifying app users based on keywords such as "pokemon go" might lead to false-positive cases. The number of tweets K can be regarded as the degree of interest for Pokémon Go. Thus, we investigated the relationship between the number of tweets and accuracy of identifying Pokémon Go users. To inspect the performance of identifying Pokémon Go users using keyword matching, we randomly sampled 150 users for each number of tweets (i.e., $K = 1, 2, 3$). Then, we manually checked whether the user is actually a Pokémon Go user or not by viewing his/her tweet timeline. There was two assessors, and the agreement was 99.3%. From this result, we considered the annotation to be reliable. Table 1 shows the results of specifying app users on the basis of keywords. From this result, we used three as the threshold for the number of tweets.

[2]https://dev.twitter.com/rest/public

	Number of tweets		
	1	2	3
Accuracy	56.7%	78.7%	84.0%

Table 1: Accuracy of identifying Pokémon Go users

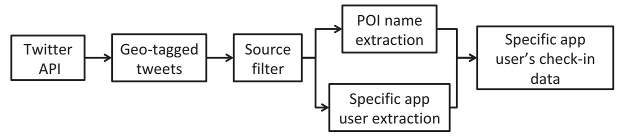


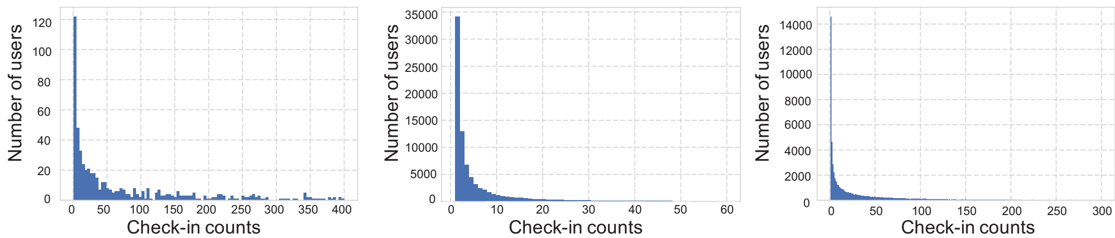
Figure 3: Flow of data collection and filtering

We collected Japanese tweets and extracted POIs in Japan from the tweets. Hence, we collected a total of 7,118,598 check-ins from 161,963 users: 154,985 check-ins from 667 Pokémon Go users, and 1,517,121 check-ins from 86,897 Instagram users. Users who used both Pokémon Go and Instagram numbered 253 and made 46,279 check-ins. Users who used both LBSs were counted in each LBS and were not excluded. These datasets were used in the experiment in Study 1 (subsequent section).

In evaluating POI predictions in Study 2, further filtering was executed in accordance with previous research (Yuan et al. 2013). First, we identified users who visited more than five locations because 20% of the check-ins were randomly deleted for the evaluation. Next, we filtered out the locations that were visited by fewer than 20 users since places where few users check-in cannot be regarded as app-specific locations. The remaining users numbered 542 for Pokémon Go and 28,971 for Instagram, and locations numbered 12,063 for Pokémon Go and 22,513 for Instagram.

Study 1: Analysis of Specific LBS User Check-in Behavior

In this subsection, we detail the experiment conducted to investigate the difference between particular LBS users and non-users. We compare specific LBS users with non-users regarding the number of check-ins, check-in locations, temporal dynamics, and the distance between successive check-ins. The former two factors have not been explored, whereas the latter two factors, which corresponded to the factors considered in the existing work, are analyzed to verify that our data indicates the same tendency as existing work.



(a) Pokémon Go (b) Instagram (c) Others

Figure 4: Number of check-ins of each LBS

	Mean	Median	SD
Pokémon Go	232.4	43.0	523.4
Instagram	17.5	2.0	107.1
Others	69.1	15.0	192.9

Table 2: Statistics of the number of check-in counts

Comparison of the number of check-ins To investigate how many users there are in each LBS user check-in, we counted the number of check-ins from the collected data. Generally, in CF, the higher the data (User-Location matrix) density is, the better the prediction performance will be (Su and Khoshgoftaar 2009). Figure 4 shows the histogram of the check-in counts. Table 2 shows the mean, median, and standard deviation of the number of check-in counts.

From these results, the User-Location matrices of Pokémon Go players and Instagram users are expected to be denser and sparser, respectively, than that of non-users.

The density of the User-Location matrix was 5.58×10^{-3} for Pokémon Go users, 5.30×10^{-4} for Instagram users, and 1.26×10^{-3} for other users. The density of the User-Location matrix of Instagram was 2.38 times sparser than that of non-users, whereas that of Pokémon Go users was 4.43 times denser than that of non-users. Therefore, the novel location predictability of Pokémon Go users is also expected to be higher than that of non-users.

Comparison of the check-in locations In this subsection, we compare the check-in location ranking of specific LBS users and non-users to investigate whether the check-in location of specific LBS users is biased or not. Table 3 lists the top 10 check-in spot rankings (determined by the number of visits) of (a) Pokémon Go, (b) Instagram, and (c) Other users. The number of visits was normalized on the basis of the maximum value.

In the ranking of Pokémon Go users, the top ranked sites include major stations such as Tokyo Station, and the tendency of Pokémon Go user completely resembles that of non-users. Meanwhile, a remarkable difference was observed in the ranking between Instagram users and non-users, specifically for "Tokyo Disneyland" and "Tokyo DisneySea." Generally, Instagram users do not remain at only a scenic place all day but go to commonly visited places such as stations. Therefore, the ranking of Instagram users consists of both commonly visited places (i.e., major stations) and places visited for LBS-specific reasons such as Tokyo Disneyland. To quantitatively compare these rankings, we calculated the Spearman rank-order correlation coefficient (Kokoska and Zwillinger 1999). The correlation coefficient was 0.333 (p-value of 0.214) between Pokémon Go users and others and -0.430 (p-value of 0.346) between Instagram users and others.

We further investigated whether a bias exists in the check-in locations of particular LBS users. To reveal the characteristics of the check-in locations of particular LBS users, we calculated the lift value (Geng and Hamilton 2006) of check-in location l by using Equation (1).

$$Lift_{U_X}(l) = \frac{p(l|U_X)}{p(l|U_Y)p(U_Y) + p(l|U_X)p(U_X)} \tag{1}$$

where U_X denotes the Pokémon Go or Instagram users, U_Y denotes non-users, and $Lift$ denotes the ratio of $p(l|U_X)$ over the average. Table 4 shows the visited sites ranked by normalized lift value. Many Pokémon Centers, which are Pokémon shops, and parks were presented in the upper rank for Pokémon Go users. The service introduction page of Pokémon Go[3], public notes that parks are good places for PokéStops. In addition, some special parks such as Tsuruma Park, are very famous as Pokémon Go locations in Japan[4]. Hence, the ranking probably reflected the influence of this phenomenon. Meanwhile, scenic sights (e.g., Mount Fuji) and historical places (e.g., Nagoya Castle) are top in the ranking of Instagram users. Thus, the ranking of Instagram users also reflected the characteristics of Instagram.

Comparison of the check-in time of a day Human mobility has an obvious temporal pattern in a day (Silva et al. 2013; Noulas et al. 2015). This corresponds to the temporal factor of novel location prediction (Yuan et al. 2013; He et al. 2016). Figure 5 shows the number of user check-ins for each time of the day. Generally, the tendency of Pokémon Go users resembles that of non-users except during late evening and nighttime (i.e., after 19:00). Meanwhile, the tendency of Instagram users is different from the non-users, particularly from early morning to before noon (03:00

[3] https://niantic.helpshift.com/a/pokemon-go/?p=web&l=en&s=pokestops&f=what-makes-a-high-quality-pokestop

[4] https://en.wikipedia.org/wiki/Tsuruma_Park

Rank	(a) Pokemon Go		(b) Instagram		(c) Others	
	Representative POI name	Rel. Freq	Representative POI name	Rel. Freq	Representative POI name	Rel. Freq
1	Tokyo Station	1	Tokyo Station	1	Akihabara Station	1
2	Shinjuku Station	0.7814	Shinjuku Station	0.7076	Tokyo Station	0.9754
3	Akihabara Station	0.776	Tokyo Disneyland	0.6323	Shinjuku Station	0.8278
4	Nagoya Station	0.6284	Tokyo DisneySea	0.6004	Shibuya Station	0.6413
5	Shibuya Station	0.5792	Shibuya Station	0.5924	Nagoya Station	0.6216
6	Ikebukuro Station	0.5792	Nagoya Station	0.5421	AKB48 Theater	0.5794
7	Osaka Station	0.5027	Akihabara Station	0.5251	Akihabara Gamers	0.5482
8	Shinagawa Station	0.4918	Club Quattro	0.4745	Ikebukuro Station	0.5134
9	Yokohama Station	0.4372	Ikebukuro Station	0.47	Shinagawa Station	0.4376
10	Animate Akihabara	0.4153	Tokyo Dome	0.4658	Haneda Airport	0.4329

Table 3: Ranking of check-in spots based on the check-in numbers of (a) Pokémon Go, (b) Instagram, and (c) Other users.

Rank	(a) Pokémon Go		(b) Instagram	
	Representative POI name	Norm. lift	Representative POI name	Norm. lift
1	Pokémon Center Tokyo DX	1	Mount Fuji	1
2	Tsuruma Park	0.645	Around Mount Fuji	0.982
3	Toda Park Station	0.5232	Yoyogi Park	0.9653
4	Tofukuji Station	0.514	Maiko Snow Resort	0.9457
5	Kinshi Park	0.4653	Naha Airport	0.9333
6	Tokyo Skytree Station	0.438	Fukuoka Airport	0.9186
7	Tsurumai Station	0.4086	Enoshima	0.9003
8	Nihonbashi	0.4019	Fukuoka Maizuru Park	0.9
9	Tochigi Stagion	0.3986	Hitachi Seaside Park	0.8837
10	Shiki Staion	0.3947	Itsukushima Shrine	0.8703
11	Tama Center Station	0.3878	Kenroku-en	0.8647
12	Kotake-mukaihara Station	0.3844	Tokyo DisneySea	0.8581
13	Pokémon Center Osaka	0.3819	Inokashira Park	0.8342
14	Otemachi Station	0.3708	Nippon Budokan	0.832
15	Pokémon Center Kyoto	0.3595	Fushimi Inari Taisha (Shrine)	0.8244
16	Warabi Station	0.3565	Nagoya Castle	0.7985
17	Kyobashi Edogrand	0.3473	Naeba Ski Resort	0.7971
18	Zushi Staion	0.3368	Meguro River Cherry Blossom	0.7771
19	Amagasaki Station	0.3253	Osanbashi Pier	0.7591
20	Pokémon Center Skytree Town	0.3207	Showa Memorial Park	0.7486

Table 4: Ranking of check-in spots based on the normalized lift value.

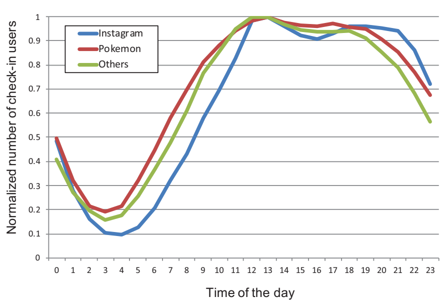


Figure 5: Number of check-ins users for time of a day

Rank	Representative POI name	norm. lift
1	Tokyo DisneySea	1
2	Ajinomoto Stadium	0.8933
3	Sensoji Temple Asakusa Kannon-Do	0.8786
4	Tokyo Disneyland	0.8442
5	Tokyo Disney Resort	0.8398
6	Kyocera Dome Osaka	0.7924
7	Tokyo Dome	0.7525
8	Shimokitazawa ERA	0.7186
9	Makuhari Messe	0.7142
10	Zepp Tokyo	0.6949

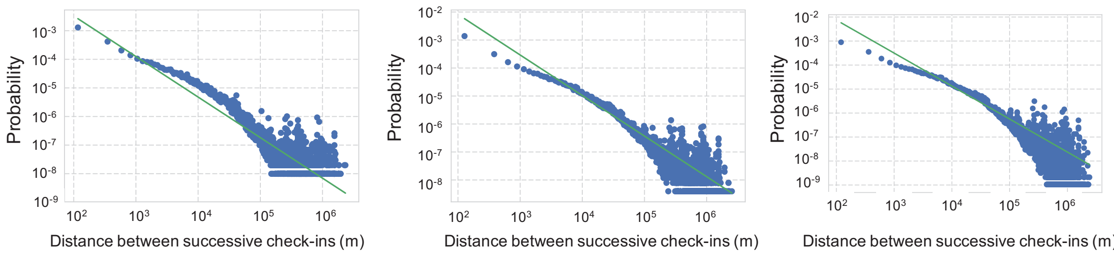
Table 5: Ranking of check-in spots based on the normalized lift value for Instagram users between 20:00 and 23:00.

to 11:00) and late evening and nighttime (after 19:00). Note that Figure 5 shows the relative number of check-ins. This means that relatively more Instagram users check-in around noon than Pokémon Go users and others. We further investigated Instagram user's check-ins in late evening and early nighttime (which is another particular characteristic of Instagram users) by calculating the lift value of check-ins from 20:00 to 23:00. Table 5 presents the results. Many large sites where events are often held in the evening and at night are present in the top ranking.

Comparison of the distance among successive check-ins
It is generally said that the distance of human mobility follows a power law (Gonzalez, Hidalgo, and Barabasi 2008; Ye et al. 2011; Yuan et al. 2013). This property was also exploited to predict a novel location of existing work (Ye et al. 2011; Yuan et al. 2013). Therefore, we investigated the distance difference among successive check-ins between specific LBS users and non-users.

First, we calculated the distance between adjacent check-ins within a day. Then, the probability of each distance was plotted (blue dot) as shown in Figure 6. The probability was calculated using *(# of samples of each distance)/(# of total samples)*. Finally, we observed that the distribution of the probability follows the power law, as shown in Figure 6.

To quantitatively compare the characteristics of each LBS, the probability distribution of successive check-ins was fitted using the power law equation. The relationship between the distance and probability is represented by the



(a) Pokémon Go (b) Instagram (c) Others

Figure 6: Distribution of distance among successive check-ins

	w_0	w_1
Pokémon Go	1.331	-1.576
Instagram	0.811	-1.448
Others	0.621	-1.374

Table 6: Parameters of the power law distribution.

following equation (Ye et al. 2011).

$$y = a \times x^b \quad (2)$$

where x is the distance between adjacent check-ins, y is the visited probability, and a and b are the parameters of the power law. By taking the logarithm on both sides of Equation (2), we can obtain the following equation.

$$\log y = w_0 + w_1 \log x \quad (3)$$

where $\log a = w_0$ and $b = w_1$. This equation is the same as that in the work of Ye et al. (Ye et al. 2011). The distribution was radically changed approximately $10^5 m$ in Figure 6. A similar tendency is also reported Yuan et al. (Yuan et al. 2013). Therefore, we fitted Equation (3) using the data between 0 and $10^5 m$. The least-squares regression was used to perform the fitting calculation.

Table 6 shows the results of the parameter estimation. The fitted line is denoted by green in Figure 6. The smaller the value of w_1 is, the nearer the tendency of users to check-in will be. As shown in Table 6, the order of the value of w_1 is $PokemonGo < Instagram < Other$. This result indicated that specific LBS users tended to check-in more in nearby locations than the non-users.

Study 1: Summary of the difference in mobility pattern

We briefly summarized the difference in mobility pattern between specific LBS users and non-users, regarding the number of check-ins, check-in locations, temporal dynamics, and the distance among successive check-ins.

(1) **Number of check-ins:** Pokémon Go players had 4.43 times more check-ins, whereas Instagram users had approximately a quarter of check-ins compared with other users.

(2) **Check-in locations:** Pokémon Go users tended to check-in at parks and nature spots/historical places more than non-users. Meanwhile, Instagram users tended to check-in at large event sites, scenic places, and historical places much more than non-users.

(3) **Check-in time of the day:** Instagram users had a different check-in time tendency from other users, especially during late evening/early night. Instagram users tended to visit large event sites in late evening and at night.

(4) **Distance among successive check-ins:** Pokémon Go and Instagram users tended to visit more nearby places than other users.

From the result of (1), the novel location predictability of Pokémon Go users is also expected to be higher than that of non-users as mentioned above. From the result of (2), we can see a remarkable difference existed between specific LBS users and non-users, especially for lift-based ranking. Thus, it is useful to predict locations for specific LBS users, and we design a method that exploits this characteristic and evaluate its effectiveness in the next section. From the results of (3) and (4), our data was validated because the results indicated the same tendency as existing work.

Study 2: Analysis of Effect on Location Prediction Using Collaborative Filtering

The results of the previous section revealed the difference in mobility patterns between specific LBS users and non-users but not how this difference influences location predictability. To this end, we performed a simple machine learning experiment to analyze the effect of specific LBS usage on location predictability.

Study 2: Methods

User-based CF is one of the most basic prediction methods. Therefore, we used the user-based CF as a basic method, and extended it to predict specific LBS users as the proposed method. In this subsection, we describe two methods to generate location predictions for specific LBS users through CF: (1) User-based CF and (2) Lift-weighted user-based CF. Our proposed lift-weighted user-based CF was based on the hypothesis that users of the same LBS are likely to visit the same location. First, we explain the location prediction by using user-based CF. Then we propose the location prediction for specific service users.

User-based Collaborative Filtering The first step of user-based CF was to calculate the similarity between the target user and other users. Then, the prediction score of a location was computed by a weighted combination of other users' check-in events. Each element $c_{i,j}$ in User-Location matrix C, which represents the visits of each user, is $c_{i,j} = 1$ when a user $u_i \in U$ checks in at location $l_j \in L$, and $c_{i,j} = 0$ when a user u_i does not check in at POI l_j. The prediction score of each location was calculated using Equation (4), which is as follows (Ye et al. 2011):

$$\hat{c}_{i,j} = \frac{\sum_{u_k \in U} w_{i,k} c_{k,j}}{\sum_{u_k} w_{i,k}} \tag{4}$$

where $w_{i,k}$ indicates the similarity weight between user u_i and user u_k.

The similarity was computed using a cosine similarity expressed by the following equation in the conventional manner of Ye et al. (Ye et al. 2011) and Yuan et al. (Yuan et al. 2013)

$$w_{i,k} = \frac{\sum_{l_j \in L} c_{i,j} c_{k,j}}{\sqrt{\sum_{l_j \in L} c_{i,j}^2}\sqrt{\sum_{l_j \in L} c_{k,j}^2}}. \tag{5}$$

Lift-weighted User-based Collaborative Filtering In this subsection, we explain how to exploit LBS usage information for location prediction. We propose a method that uses the lift value to add weight to an LBS-specific location in the general CF framework. Here, we introduce the simplest method that incorporates the lift value into CF because our main purpose is to validate our hypothesis that we can improve the performance of novel location prediction by considering whether the target user uses a specific LBS or not. The computational cost of the proposed method in addition to the user-based CF is low because the proposed method can be calculated by adding one multiplication to the user-based CF thanks to the simplicity.

Based on the analysis of study 1, users who use the same LBS are likely to visit similar places. Thus, the prediction score of lift-weighted CF $\hat{c}_{i,j}^{(s)}$ of each location for the user u_i was calculated by using the following equation.

$$\hat{c}_{i,j}^{(s)} = (1 + Lift(l_j)) \times \frac{\sum_{u_k \in U_X} w_{i,k} c_{k,j}}{\sum_{u_k \in U_X} w_{i,k}} \tag{6}$$

To reduce the computational cost, we filtered the set of users using only specific LBS users U_X. $w_{i,k}$ was calculated using Equation (5).

We calculate the prediction scores of each location for all test data using each model. A ranked list is formed by sorting all locations in accordance with their prediction scores, and top-N scored locations are recommended.

Study 2: Results

Evaluation Settings In our experiment, we randomly selected 20% of users' visited locations as test data (i.e., these locations are regarded as yet to be visited) and predicted locations from the remaining 80% of check-ins for each user. Thereafter, we used the top-N scored locations as predictions and varied the parameter as N=1, 5, 10, 20. To increase the robustness of the evaluation, we evaluated the prediction performance five times by changing the random seed for Pokémon Go users and calculated the average of these results because the number of Pokémon Go users is limited.

Metrics We use precision@N, recall@N, and accuracy@N as evaluation metrics, following previous studies on location recommendation (Ye et al. 2011; Yuan et al. 2013; Kurashima et al. 2013). Precision@N is the ratio of the recovered locations to the top-N predicted locations, recall@N is the ratio of the recovered locations to the set of locations selected randomly, and accuracy@N is the percentage of users who had least one location in their ground truth of the top-N predicted locations.

Comparison Methods We compared the proposed lift-weighted user-based CF method with two baseline methods. **User-based CF:** The first baseline method is user-based CF described in Eq. (4) because it has been widely used as the most basic method in many existing works (Ye et al. 2011; Yuan et al. 2013).
User-based CF + Geo: The second baseline method exploits the geographical factor by a power law proposed by Yuan et al. (Yuan et al. 2013). We selected this geographical-factor-based method for comparison because the geographical factor for recommendation is still underexplored in recent research (Wang et al. 2018). In addition, it is one of the most basic methods. We calculated the check-in probability that a user will check-in at location l_j after l_i by the following conditional probability. This equation follows the work of Yuan et al. (Yuan et al. 2013).

$$p(l_j|l_i) = \frac{a \times x(l_i, l_j)^b}{\sum_{l_k \in L, l_k \neq l_i} a \times x(l_i, l_k)^b} \tag{7}$$

where $x(l_i, l_j)$ denotes the distance between the location l_i and l_j. Given a user u and the check-in history of the user h_u, we can calculate the score that is proportional to the check-in probability of each location l as follows.

$$\hat{c}_{i,j}^{(g)} = P(l)P(l|h_u) \quad \propto \quad P(l)P(h_u|l) \tag{8}$$

$$= \quad P(l) \prod_{l' \in h_u} p(l'|l) \tag{9}$$

where $P(l)$ denotes the prior probability that is proportional to the ratio of users checked in l out of all users in the dataset. We used the Bayes rule in Equation (8). After the above score based on the geographical factor is calculated, we integrated the geographical-factor-based score calculated in Equation (9) and the score of user-based CF by linear combination. This combined score is referred to as User-based CF + Geo. These two methods have different score scales, thus we normalize each score as follows.

$$\tilde{c}_{i,j} = \frac{\hat{c}_{i,j}}{\max_{l_j \in L - h_u}(\hat{c}_{i,j})} \tag{10}$$

$$\tilde{c}_{i,j}^{(g)} = \frac{\hat{c}_{i,j}^{(g)}}{\max_{l_j \in L - h_u}(\hat{c}_{i,j}^{(g)})} \tag{11}$$

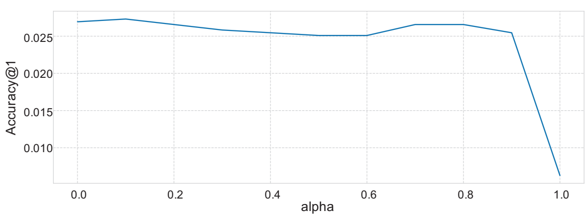


Figure 7: The results of different parameters for User-based CF + Geo

Then, we computed the integrated score as follows.

$$c_{i,j} = (1-\alpha) \times \tilde{c}_{i,j} + \alpha \times \tilde{c}_{i,j}^{(g)} \quad (12)$$

where α is the tuning parameter. The reason we combined the score of the geographical factor and user-based CF is to compare each method under fair conditions because the proposed lift-weighted user-based CF method contains the calculation of the user-based CF. To tune the parameter, we varied α from 0.0 to 1.0 in 0.1 steps regarding accuracy@1. Figure 7 shows the results. The performance curve peaks around $\alpha = 0.1$. Thus, we set $\alpha = 0.1$ in comparison. We compared the user-based CF + Geo method only for the case of Pokémon Go users because its computational cost, which depends on the number of users and pairs of locations, is extremely high for the case of Instagram users.

Evaluation Results For a quantitative evaluation, we compared the effectiveness of the proposed lift-weighted user-based CF method with that of the user-based CF (baseline method). Tables 7 and 8 show the evaluation results for Pokémon Go and Instagram users, respectively. Scores in bold in Table 7 are the best results. The proposed method outperformed user-based CF except for precision@1, recall@1, and accuracy@1 for Pokémon Go users. The two-tailed t-test showed that the difference was significant for all cases for Instagram users. The level of statistical significance was 0.999. Meanwhile, for Pokémon Go users, † and ‡ indicate statistical significance at the 0.95 and 0.995 level with respect to both User-based CF and User-based CF + Geo, respectively.

Study 2: Discussion

In Study 2, our hypothesis was that the accuracy of novel location prediction can be improved by considering the use of specific LBSs. Results in Tables 7 and 8 support our hypothesis. Here, we discuss these results in more depth. Notably, the performance of novel location prediction was higher for Instagram users than for Pokémon Go users despite the User-Location matrix of Instagram being sparser than that of Pokémon Go users and other users. We suppose that there are two reasons for this: (1) the qualitative difference in the check-in data on Foursquare and Instagram, and (2) the difference between digital content and real-world content. For the first reason, the check-in data of Pokémon Go is not only the data for when the user plays Pokémon Go but also does other activities, whereas the check-in data of Instagram is exactly for when the user uses Instagram. Therefore, the bias of the location of Instagram users may be more clearly reflected than that of Pokémon Go users. The second reason is that the content distribution of digital content is easier to change than that of real-world content. That is, the content of Pokémon Go can be easily changed by service providers (i.e., Niantic). To investigate the difference between digital content and real-world content, we calculated the top 100 rankings of check-in spots on the basis of the lift value for each month (from January to June). Then we compared the number of months in which the locations appeared in the ranking as shown in Figure 8. There is an apparent difference between Pokémon Go and Instagram. Approximately 24.8% of the locations appeared in the top 100 rankings during six months for Instagram. On the other hand, fewer locations (18.5%) appeared in the top 100 rankings during six months for Pokémon Go. In addition, 28.4% of the locations appeared in the ranking in one month for Instagram, whereas 37.1% of the locations appeared in the ranking in one month for Pokémon Go. We suppose that this phenomenon reflects the difference in the characteristics of digital content and real-world content. From this result, it may be more effective for Pokémon Go users to use time-varying lift value (i.e., calculating the lift value for short term, such as month by month).

Implications for Location-based Services

Our results lead to implications for LBSs. One LBS application is in intelligent personal assistants such as Apple's Siri (Apple 2017) and Google Now (Google 2017). Traffic and weather information are provided on the basis of the current or predicted location of a user. The locations are mainly regular places, such as home, workplace, and a few frequently visited places. However, information for novel places is also important for a user. Hence, we suggest providing location-related information of novel places on the basis of novel location prediction. We consider the essential implications of the study is that we can leverage the results of novel location prediction which is more accurate for LBS applications. From this viewpoint, we may also apply our results to location-based advertisements as well as information provision. In addition, we may exploit the results of novel location prediction for public health because human mobility affects the spread of the infection (Meloni et al. 2011). Another application is to support the prevention of smartphone overuse. Recently, smartphone makers such as Apple are providing a function to restrict smartphone use (e.g., Apple's screen time[5]), because overuse of smartphones is a problem. The fact that LBS usage information helps novel location prediction may be conversely said to prevent a user from overusing his/her smartphone (i.e., specific LBSs) too much, especially for a user who is correctly estimated with high probability. Thus, our analysis may help prevent users from overusing their smartphones.

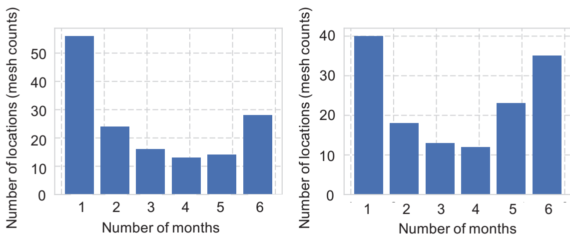
[5]https://www.apple.com/ios/ios-12/

	Precision				Recall				Accuracy			
	pre@1	pre@5	pre@10	pre@20	rec@1	rec@5	rec@10	rec@20	acc@1	acc@5	acc@10	acc@20
User-based CF	0.0269	0.0234	0.0221	0.0218	0.0004	0.0010	0.0018	0.0029	0.0269	0.1033	0.1827	0.3107
User-based CF + Geo	**0.0273**	0.0231	0.0222	0.0217	0.0004	0.0010	0.0018	0.0029	**0.0273**	0.1015	0.1845	0.3070
Lift-weighted CF	0.0244	**0.0263**‡	**0.0251**‡	**0.0236**†	0.0004	**0.0012**‡	**0.0020**‡	**0.0032**†	0.0244	**0.1221**‡	**0.2000**†	**0.3229**

Table 7: Prediction performance for Pokémon Go users

	Precision				Recall				Accuracy			
	pre@1	pre@5	pre@10	pre@20	rec@1	rec@5	rec@10	rec@20	acc@1	acc@5	acc@10	acc@20
User-based CF	0.0223	0.0183	0.0168	0.0151	0.0017	0.0041	0.0067	0.0101	0.0223	0.0838	0.1457	0.2370
Lift-weighted CF	0.0286	0.0234	0.0205	0.0176	0.0024	0.0059	0.0092	0.0133	0.0286	0.1059	0.1683	0.2568

Table 8: Prediction performance for Instagram users



(a) Pokémon Go (b) Instagram

Figure 8: Comparison of the number of locations ranked in top 100.

Limitations

Our work has the following limitations.
(1) **Data source of human mobility**. Because we use the Foursquare and Instagram's check-in data for our experiment, a possibility exists that we cannot obtain all the visited locations. In addition, the user might not actually check-in at the location. However, because check-in data is actively provided by the user, our analysis can consider his/her privacy. In addition, regular places were believed to be checked-in at less by the user than novel places. Hence, these characteristics may be suitable for our analysis. Besides coverage, there may be a difference between the demographic distribution of an actual service (i.e., Pokémon Go and Instagram) and the data collected through Twitter. One alternative method to record a user's location is always logging the location of a user by using GPS. However, in this case, the visited POI needs to be estimated from GPS coordinates, which may induce error because visited POI estimation from GPS logs is underexplored (Nishida et al. 2014). On the other hand, the advantage of using the check-in data is that we can exactly obtain the visited POI because the check-in data is actively provided by the user. Therefore, we used the check-in data collected from Twitter.
(2) **Trends in the popularity of each LBS**. Each LBS has a popularity trend. If an LBS is very popular at certain period, the size of the effect becomes large and vice versa. Thus, we have to consider the dynamic property of LBSs.
(3) **Causality**. Because our analysis is not causal inference but association analysis, we cannot conclude that the LBSs are the cause of location bias. However, if we can confirm the causality, we can leverage the causality of mobile app usage and visited locations to online to offline (O2O) marketing, because if a user installs a specific app, then that the user may possibly go to a specific location.
(4) **Scale of dataset**. Although our dataset contains at least 150,000 check-ins for Pokémon Go users, the number of the users is limited. We can expand the number of Pokémon Go users by simply extending data collection period. However, this cannot solve the problem, because the check-in locations of Pokémon Go users have time-varying characteristics as discussed in Study 2. In addition, we considered that even though the dataset is not sufficiently large, it is not too small because Colley *et al.* (Colley et al. 2017) collected 375 Pokémon users from all over the world, whereas we collected 667 users with almost 150,000 check-ins within Japan. In addition to the size of the dataset, because our dataset was more focused to specific region, our dataset was more suitable for the analysis of check-in behavior.

Conclusion

In this study, we analyzed the influence of location-based services (LBSs) on human mobility patterns and exploited it for location prediction. In the first part of this study, we investigated the difference between specific LBS users and non-users with regard to the number of check-ins, check-in locations, temporal dynamics, and the distance among successive check-ins. We found a remarkable difference in mobility between specific LBS users and non-users (e.g., check-in locations). These insights are beneficial to understand human behavior and propose novel methods. Thereafter, we proposed a lift-weighted user-based collaborative filtering (CF) that exploits the characteristics of the bias of check-in locations to predict novel location, and we analyzed the impact of LBS on location predictability. We compared conventional user-based CF with lift-weighted user-based CF to predict the locations for each user. In the lift-weighted user-based CF, we extended the method of user-based CF by weighing the lift value. Foursquare and Instagram check-in data was used to evaluate the effect for location predictability. The result showed that the novel location predictability of Instagram users is up to relatively 43.9% higher than that of non-users in terms of recall.

In future work, we would like to introduce a unified framework of novel location prediction that exploits all factors investigated in Study 1 with the help of LBS information. In addition, we will evaluate the movement effects of users of LBSs other than Pokemon Go and Instagram.

References

Althoff, T.; White, R. W.; and Horvitz, E. 2016. Influence of pokémon go on physical activity: Study and implications. *Journal of Medical Internet Research* 18(12).

Apple. 2017. ios - siri, apple. https://www.apple.com/ios/siri/.

Boim, R.; Milo, T.; and Novgorodov, S. 2011. Diversification and refinement in collaborative filtering recommender. In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*, CIKM '11, 739–744. New York, NY, USA: Association for Computing Machinery.

Cho, E.; Myers, S. A.; and Leskovec, J. 2011. Friendship and mobility: User movement in location-based social networks. In *Proceedings of KDD '11*, 1082–1090.

Colley, A.; Thebault-Spieker, J.; Lin, A. Y.; Degraen, D.; Fischman, B.; Häkkilä, J.; Kuehl, K.; Nisi, V.; Nunes, N. J.; Wenig, N.; Wenig, D.; Hecht, B.; and Schöning, J. 2017. The geography of pokémon go: Beneficial and problematic effects on places and movement. In *Proceedings of CHI '17*, 1179–1192.

Geng, L., and Hamilton, H. J. 2006. Interestingness measures for data mining: A survey. *ACM Computing Surveys* 38(3).

Geohash. 2008. Geohash. https://twitter.com/.

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

Google. 2017. Google apps. https://www.google.com/search/about/.

He, J.; Li, X.; Liao, L.; Song, D.; and Cheung, W. K. 2016. Inferring a personalized next point-of-interest recommendation model with latent behavior patterns. In *Proceedings of AAAI'16*, 137–143. AAAI Press.

Ho, Y.-C.; Chiang, Y.-T.; and Hsu, J. Y.-J. 2014. Who likes it more? mining worth-recommending items from long tails by modeling relative preference. In *Proceedings of the 7th ACM International Conference on Web Search and Data Mining*, WSDM '14, 253–262. New York, NY, USA: Association for Computing Machinery.

Japanese Statistics Bureau. 1996. Grid square statistics. http://www.stat.go.jp/english/data/mesh/05.html.

Kaminskas, M., and Bridge, D. 2016. Diversity, serendipity, novelty, and coverage: A survey and empirical analysis of beyond-accuracy objectives in recommender systems. *ACM Trans. Interact. Intell. Syst.* 7(1):2:1–2:42.

Kokoska, S., and Zwillinger, D. 1999. *CRC standard probability and statistics tables and formulae*. Crc Press.

Kunaver, M., and Požrl, T. 2017. Diversity in recommender systems–a survey. *Knowledge-Based Systems* 123:154–162.

Kurashima, T.; Iwata, T.; Hoshide, T.; Takaya, N.; and Fujimura, K. 2013. Geo topic model: Joint modeling of user's activity area and interests for location recommendation. In *Proceedings of WSDM '13*, 375–384.

Li, Z.; Ding, B.; Han, J.; Kays, R.; and Nye, P. 2010. Mining periodic behaviors for moving objects. In *Proceedings of KDD '10*, 1099–1108.

Li, H.; Ge, Y.; Hong, R.; and Zhu, H. 2016. Point-of-interest recommendations: Learning potential check-ins from friends. In *Proceedings of KDD '16*, 975–984.

Lian, D.; Xie, X.; Zheng, V. W.; Yuan, N. J.; Zhang, F.; and Chen, E. 2015. Cepr: A collaborative exploration and periodically returning model for location prediction. *ACM Trans. Intell. Syst. Technol.* 6(1):8:1–8:27.

Lian, D.; Zheng, V. W.; and Xie, X. 2013. Collaborative filtering meets next check-in location prediction. In *Proceedings of WWW '13 Companion*, 231–232.

Ma, H.; King, I.; and Lyu, M. R. 2009. Learning to recommend with social trust ensemble. In *Proceedings of SIGIR '09*, 203–210.

Meloni, S.; Perra, N.; Arenas, A.; Gómez, S.; Moreno, Y.; and Vespignani, A. 2011. Modeling human mobility responses to the large-scale spreading of infectious diseases. *Scientific reports* 1:62.

Nishida, K.; Toda, H.; Kurashima, T.; and Suhara, Y. 2014. Probabilistic identification of visited point-of-interest for personalized automatic check-in. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '14, 631–642. New York, NY, USA: ACM.

Noulas, A.; Scellato, S.; Lambiotte, R.; Pontil, M.; and Mascolo, C. 2012. A tale of many cities: universal patterns in human urban mobility. *PloS one* 7(5):e37027.

Noulas, A.; Shaw, B.; Lambiotte, R.; and Mascolo, C. 2015. Topological properties and temporal dynamics of place networks in urban environments. In *Proceedings of WWW '15 Companion*, 431–441. New York, NY, USA: ACM.

Paavilainen, J.; Korhonen, H.; Alha, K.; Stenros, J.; Koskinen, E.; and Mayra, F. 2017. The pokémon go experience: A location-based augmented reality mobile game goes mainstream. In *Proceedings of CHI '17*, 2493–2498.

Sheldon, P., and Bryant, K. 2016. Instagram: Motives for its use and relationship to narcissism and contextual age. *Computers in human Behavior* 58:89–97.

Silva, T. H.; Vaz de Melo, P. O. S.; Almeida, J. M.; Salles, J.; and Loureiro, A. A. F. 2013. A comparison of foursquare and instagram to the study of city dynamics and urban social behavior. In *Proceedings of UrbComp '13*, 4:1–4:8. New York, NY, USA: ACM.

Song, C.; Qu, Z.; Blumm, N.; and Barabási, A.-L. 2010. Limits of predictability in human mobility. *Science* 327(5968):1018–1021.

Su, X., and Khoshgoftaar, T. M. 2009. A survey of collaborative filtering techniques. *Advances in artificial intelligence* 2009:4.

Tasse, D.; Liu, Z.; Sciuto, A.; and Hong, J. I. 2017. State of the geotags: Motivations and recent changes. In *Proceedings of ICWSM '17*, 538–541.

von Watzdorf, S., and Michahelles, F. 2010. Accuracy of positioning data on smartphones. In *Proceedings of LocWeb '10*, 2:1–2:4. New York, NY, USA: ACM.

Wang, Y.; Yuan, N. J.; Lian, D.; Xu, L.; Xie, X.; Chen, E.; and Rui, Y. 2015. Regularity and conformity: Location prediction using heterogeneous mobility data. In *Proceedings of KDD '15*, 1275–1284.

Wang, J.; Li, M.; Han, J.; and Wang, X. 2016. Modeling check-in preferences with multidimensional knowledge: A minimax entropy approach. In *Proceedings of WSDM '16*, 297–306. New York, NY, USA: ACM.

Wang, H.; Shen, H.; Ouyang, W.; and Cheng, X. 2018. Exploiting poi-specific geographical influence for point-of-interest recommendation. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, IJCAI'18, 3877–3883. AAAI Press.

Ye, M.; Yin, P.; Lee, W.-C.; and Lee, D.-L. 2011. Exploiting geographical influence for collaborative point-of-interest recommendation. In *Proceedings of SIGIR '11*, 325–334.

Yu, Y., and Chen, X. 2015. A survey of point-of-interest recommendation in location-based social networks. In *Workshops at the Twenty-Ninth AAAI Conference on Artificial Intelligence*, volume 130.

Yuan, Q.; Cong, G.; Ma, Z.; Sun, A.; and Thalmann, N. M. 2013. Time-aware point-of-interest recommendation. In *Proceedings of SIGIR '13*, 363–372.

Zhang, C.; Liang, H.; and Wang, K. 2016. Trip recommendation meets real-world constraints: Poi availability, diversity, and traveling time uncertainty. *ACM Trans. Inf. Syst.* 35(1).