CEPR: A Collaborative Exploration and Periodically Returning Model for Location Prediction

DEFU LIAN, University of Science and Technology of China
XING XIE, Microsoft Research
VINCENT W. ZHENG, Advanced Digital Sciences Center, Singapore
NICHOLAS JING YUAN, Microsoft Research
FUZHENG ZHANG and ENHONG CHEN, University of Science and Technology of China

With the growing popularity of location-based social networks, numerous location visiting records (e.g., check-ins) continue to accumulate over time. The more these records are collected, the better we can understand users' mobility patterns and the more accurately we can predict their future locations. However, due to the personality trait of neophilia, people also show propensities of novelty seeking in human mobility, that is, exploring unvisited but tailored locations for them to visit. As such, the existing prediction algorithms, mainly relying on regular mobility patterns, face severe challenges because such behavior is beyond the reach of regularity. As a matter of fact, the prediction of this behavior not only relies on the forecast of novelty-seeking tendency but also depends on how to determine unvisited candidate locations. To this end, we put forward a Collaborative Exploration and Periodically Returning model (CEPR), based on a novel problem, Exploration Prediction (EP), which forecasts whether people will seek unvisited locations to visit, in the following. When people are predicted to do exploration, a state-of-the-art recommendation algorithm, armed with collaborative social knowledge and assisted by geographical influence, will be applied for seeking the suitable candidates; otherwise, a traditional prediction algorithm, incorporating both regularity and the Markov model, will be put into use for figuring out the most possible locations to visit. We then perform case studies on check-ins and evaluate them on two large-scale check-in datasets with 6M and 36M records, respectively. The evaluation results show that EP achieves a roughly 20% classification error rate on both datasets, greatly outperforming the baselines, and that CEPR improves performances by as much as 30% compared to the traditional location prediction algorithms.

Categories and Subject Descriptors: H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval—*Clustering, information filtering*; H.2.8 [**Database Management**]: Database Applications—*Data mining, spatial databases and GIS*

General Terms: Algorithms, Design, Experimentation

Additional Key Words and Phrases: Location-based services, Exploration Prediction, social network, location prediction, location recommendation

---

This research was partially supported by a grant from the National Science Foundation for Distinguished Young Scholars of China (Grant No. 61325010).
Authors' addresses: D. Lian, Big Data Research Center, University of Electronic Science and Technology of China, No. 2006 Xiyuan Avenue, West Hi-tech Zone, Chengdu, China; email: dove@uestc.edu.cn; X. Xie and N. J. Yuan, Microsoft Research, Building 2, No. 5 Danling Street, Haidian District, Beijing, China; emails: {xingx, nicholas.yuan}@microsoft.com; V. W. Zheng, Advanced Digital Sciences Center, 1 Fusionopolis Way, #08-10, Connexis North Tower, Singapore; email: vincent.zheng@adsc.com.sg; F. Zhang and E. Chen, University of Science and Technology of China, No. 443 Huangshan road, Hefei, China; emails: zhfzh@mail.ustc.edu.cn, cheneh@ustc.edu.cn.

DOI: http://dx.doi.org.10.1145/2629557

ACM Reference Format:
Defu Lian, Xing Xie, Vincent W. Zheng, Nicholas Jing Yuan, Fuzheng Zhang, and Enhong Chen. 2015. CEPR: A collaborative exploration and periodically returning model for location prediction. ACM Trans. Intell. Syst. Technol. 6, 1, Article 8 (April 2015), 27 pages.
DOI: http://dx.doi.org.10.1145/2629557

1. INTRODUCTION

With the proliferation of smartphones and the development of positioning technologies, location information can be acquired more easily than ever before. This development has triggered a new kind of social network service—location-based social networks (LBSNs), such as Foursquare, Jiepang, Gowalla, and so on. In an LBSN, people can not only track and share location-related information of an individual but also leverage collaborative social knowledge learned from others. As LBSNs have been increasing in popularity in recent years, numerous location-visiting records, for example, check-ins (being used to represent the process of announcing and sharing users' current locations), continue to accumulate over time. The more these records are collected, the better we can understand users' mobility patterns and the more accurately we can predict their future locations.

Accurate location prediction plays an important part in urban planning, traffic forecasting, advertising, and recommendations; thus, it has recently attracted a lot of attention [Ashbrook and Starner 2002; Song et al. 2004, 2010b; Cho et al. 2011; Chang and Sun 2011; Etter et al. 2012; Gao et al. 2012a; Noulas et al. 2012a]. Most researchers in this field have figured out the extreme importance of individual history in location prediction, so their proposed prediction algorithms have heavily depended on regular or repetitive mobility patterns. However, due to the personality trait of neophilia, which is related to the dopamine system [Ebstein et al. 1996], people also show propensities of novelty seeking in human mobility, that is, exploring unvisited but tailored locations for them to visit. For example, people may try some novel restaurants to have a change in diet. In fact, such behavior is more salient on LBSNs due to users' proactive check-in behavior. Specifically, according to simple statistics on a Gowalla check-in dataset from Cho et al. [2011], on average, users log more than 35% check-ins at novel (i.e., previously unchecked-in) locations each day even after half a year; given that 80% of days of check-in history of each user is observed, check-ins at locations that are not in the observed history of each user occupy over 50% of the rest check-ins. Due to the existence of such behavior and beyond the reach of regularity, those existing prediction algorithms will face severe challenges. As a matter of fact, the prediction of this behavior not only relies on the forecast of novelty-seeking tendency but also depends on how to determine those unvisited candidate locations.

To achieve the goal of forecasting novelty-seeking tendency, we put forward an *Exploration Prediction* (EP) problem, which forecasts whether people will seek unvisited locations to visit, in the following. More specifically, EP would like to predict whether a user's next visiting location exists in his or her location history and is thus naturally boiled down to a binary classification problem; it should depend on not only the personality trait of novelty seeking (e.g., how much a user loves exploring novel locations) but also current status of neophilia (e.g., whether a user is doing exploration now and how many opportunities are left for a user to seek novel locations). For the latter case about current neophilia status, assuming that a user has visited lots of locations near his or her activity areas, the propensity of seeking novel locations should be smaller. This is in line with the assumption in Song et al. [2010a] that the probability of visiting a novel location next is in proportion to $S^{-\gamma}$, where S is the distinct number of locations and $\gamma > 0$ is a parameter to control the exploration tendency. Moreover, providing context information, EP should also be time dependent and location dependent. For example, users usually have distinct degrees of novelty-seeking tendency at different times; for example, users may prefer to explore during weekends. If a user has arrived in an unfamiliar location (e.g., city), his or her propensity of novelty seeking will increase. In order to deeply understand this problem and differentiate the contribution of these factors, we conduct case studies on check-ins from location-based social networks and evaluate it on two large-scale check-in datasets with 6M and 36M check-ins, respectively. The results show that EP can achieve around a 20% classification error rate on both datasets and greatly outperform a MostFrequent classifier, which makes an Exploration Prediction according to the frequency of individual novelty seeking, with a 42% and 49% error rate on Gowalla and Jiepang, respectively.

Another goal to cope with novelty seeking in prediction is finding those unvisited candidate locations. Someone may consider that people usually explore popular locations, so it may be good to aggregate all users' history to get a popularity-based model. However, its contribution in finding novel candidates is small compared to other factors in novel location prediction [Noulas et al. 2012b; Gao et al. 2012b] or location recommendation [Zheng et al. 2010a, 2010b, 2011; Ye et al. 2011; Cheng et al. 2012] due to lack of personalization. In this case, another one may wonder, why not directly apply recommendation techniques for location prediction, no matter whether the next one is novel or not, similar to Lian et al. [2013]? Unfortunately, these solutions don't work well, because, according to our observations, most of the existing location recommendation algorithms tend to find unvisited but tailored locations for users, and it is difficult for them to capture regularity in human mobility. If we design new recommendation algorithms to model regularity directly, their capacity of finding novel and suitable locations for users becomes weaker, since not only may these models not well estimate the similarity between users due to the scarcity of mobility data, but also the recommending locations for users are dominated by the regular ones of their similar users. The latter reason will greatly affect location recommendation. For example, if a user A regularly travels between her residence and workplace and occasionally goes to restaurants x and y, and another user B also regularly goes back and forth between her own residence and workplace (different from user A) and has been to restaurants x and z at times, the collaborative filtering algorithms taking regularity into consideration tend to recommend the residence and workplace of user B rather than restaurant z to user A, which may be difficult to accept for the user A.

To this end, we propose a *Collaborative Exploration and Periodically Returning* model (CEPR) based on the Exploration Prediction problem to jointly make the best use of both recommendation models and prediction models. When people are predicted to do exploration, the state-of-the-art recommendation algorithm [Ye et al. 2011; Noulas et al. 2012b], armed with collaborative social knowledge and assisted by geographical influence, will be applied for seeking suitable candidates; otherwise, a traditional prediction algorithm, which incorporates both regularity and a Markov model, will be put into use for figuring out the most possible locations to visit. When EP outputs an exploration tendency instead of fully accurate prediction results, their outputs can be *interpolated* together.

We perform the case studies of the CEPR model on check-ins. However, we note that in the state-of-the-art location recommendation algorithms on check-ins, there exist two drawbacks. First, inferring geographical influence for POIs away from a user's activity areas is unnecessary since their influence is usually small so that it is difficult for them to show up in the user's candidates' list. Second, the integration of collaborative social knowledge (including user preference and social influence) with geographical influence is tuned manually so that it requires much human effort. To address the first flaw, we directly perform two-dimensional kernel density estimation (KDE) to infer geographical influence instead of making an assumption of power law [Ye et al. 2011] or Gaussian Mixture [Cheng et al. 2012] since it also meets Tobler's First Law of Geography [Tobler 1970]. And KDE can be completed by a low time/cost influence propagation approach under the guarantee of approximating recommendation performance. To overcome the second defect, we leverage a learning-to-rank algorithm [Burges et al. 2005; Liu 2009], taking them as input to learn a final representation of scoring function. In addition to figuring out the drawbacks of recommendation algorithms in the CEPR model, we also observe the problems of traditional prediction algorithms stemming from the scarcity of individual check-ins and the large number of each user's visited locations. Therefore, we exploit kernel smoothing techniques on time distribution at a given location to model regularity and leverage the widely used interpolation techniques of a high-order Markov model with the low-order ones in language models to estimate a Markov model. Then, regularity and a Markov model are seamlessly incorporated into a hidden Markov model framework for location prediction when we consider locations of check-ins as hidden states and other information as observations. After the study of the CEPR model on check-ins, we perform extensive evaluations on two aforementioned check-in datasets. The experimental results indicate that CEPR can improve prediction performance by as much as 30% compared to the traditional prediction algorithms.

In summary, the contributions of this article are defined by the following points.

(1) We propose an Exploration Prediction problem, which forecasts whether a user's next visit is an exploration to an unvisited location, and boil it down to a binary classification problem. We further conduct case studies on check-ins and evaluate it on two large-scale check-in datasets with 6M and 36M check-ins, respectively. The results show that EP can achieve around a 20% classification error rate and greatly outperforms the baselines on both datasets.
(2) We propose a Collaborative Exploration and Periodically Return model based on EP, which integrates location recommendation and prediction algorithms according to the results from Exploration Prediction. We then evaluate it on two aforementioned check-in datasets. The results show that CEPR can be 30% more effective compared to the traditional location prediction algorithms on both datasets.
(3) To cope with the challenge stemming from the scarcity of individual check-ins and the large number of each user's visited locations, we exploit kernel smoothing techniques on time distribution at a given location to model regularity, and leverage the widely used interpolation techniques of a high-order Markov model with the low-order ones in language models to estimate a Markov model. Both of them are further incorporated into an HMM framework.
(4) To address the drawbacks of existing check-in location recommendation algorithms, we make several improvements, including reducing the time complexity of computing geographical influence and applying a learning-to-rank algorithm to integrate those important factors.

The rest of this article is organized as follows. We first survey related work in Section 2. Since our investigated location-visiting records are check-ins from LBSNs, we then analyze two check-in datasets in Section 3 to figure out the special characteristics of check-ins. After that, we put forward the Collaborative Exploration and Periodically Returning model based on the Exploration Prediction problem and conduct case studies on check-ins in Section 4, where we also present a check-in location recommendation and prediction algorithm. After presenting models, in Section 5, we evaluate EP and CEPR on two check-in datasets. Following this, we draw conclusions and discuss future work in Section 6.

2. RELATED WORK

Location prediction has been studied for a long time. In the past, the concentration has been on mining frequent mobility patterns from the GPS trajectories. Ashbrook and Starner [2002] presented a system that automatically clustered the GPS data taken over an extended period of time into meaningful locations and incorporated them into a Markov model to predict future movement. Song et al. [2004] reported the results of the empirical evaluation of location predictors on WiFi mobility data and concluded that a second-order Markov model performed the best. Monreale et al. [2009] built a decision-tree-based prediction system to learn from trajectory patterns and made location predictions by means of finding the best matching path in the tree. Gonzalez et al. [2008] analyzed mobile phone data to understand users' individual mobility patterns and demonstrated that human trajectories show a high degree of temporal and spatial regularity. To account for the empirically observed scaling laws in human mobility, Song et al. [2010a] introduced two principles similar to CEPR to build a microscopic model for individual human mobility. However, with the aim of predicting most of the pertinent scaling exponents, in this model, people explore novel locations based on the distance distribution between consecutive locations and return to regular locations with probabilities in proportion to the visiting frequency.

In addition to the study of location prediction on mobility trajectories, location-based social networks have recently become a hot topic of research. It has encountered new opportunities and challenges because users' mobility data continues to accumulate over time and more information comes with this. One of them is the introduction of social networks so that users' mobility may be directly or indirectly influenced by their friends. Therefore, there is much work to be done on mobility modeling with the presence of social relationships. For example, Cho et al. [2011] proposed a periodical and social-based model to predict the next location and concluded a small but significant effect of social relationships. Gao et al. [2012a, 2013] first built a Hierarchical Pitman-Yor (HPY) process model for each user to capture the long-range dependence among locations while at the same time of producing the power law distribution of check-in frequency, which resembles the observations in the check-in behavior [Cheng et al. 2011; Gao et al. 2012a]. They also took social relationships into account to build a hybrid model so as to integrate the prediction of users' self with friends and also discovered a small impact of social relationships on location prediction. Sadilek et al. [2012] proposed a Dynamic Bayesian Network model to predict users' future locations based on their friends' with the presence of temporal information. Noulas et al. [2012a] and Chang and Sun [2011] built a prediction model using feature engineering and took into consideration plenty of features, including location popularity, users' self and friends' preference, topics and categories of locations, and so on. Nevertheless, the major discovery from their experiments is still the importance of users' preferences. This is because users often return to previously visited locations, in particular after a certain period of the service usage [Cheng et al. 2011]. Although location prediction on LBSNs has exploited the influence of friends, which can come from human collective behaviors (e.g., having dinner with friends) or the word-of-mouth recommendation from them, it doesn't fully explore the collaborative social knowledge. For example, it should at least include the knowledge from users sharing similar mobility patterns. Therefore, the proposed CEPR model is different from these existing approaches. It not only tries to fully capture collaborative social knowledge based on recommendation techniques but also makes better use of the individual power of regularity and recommendation based on Exploration Prediction so that it prevents regularity (individual preference) from always playing a dominating part in location prediction.

To fully exploit the collaborative social knowledge, many researchers took out location exploration history and concentrate on the prediction on them. Since this task cannot be finished by traditional prediction algorithms, it usually resorts to recommendation techniques. Ye et al. [2011], Gao et al. [2012b], and Noulas et al. [2012b] employed collaborative filtering models, leveraging the similarity between users on mobility patterns and social relationships, for POI recommendations. Yang et al. [2013] and Liu and

Xiong [2013] enhanced POI recommendation with textual information, such as tips and categories, of POIs. In addition to leveraging the extra information for alleviating the data scarcity problem, Zhang et al. [2013] proposed localized matrix factorization for recommendation based on matrix block diagonal forms. And of note is Ye et al. [2011], who exploited geographical influence for location recommendation by assuming the power law distribution of distance between pairs of locations. Hence, the POIs nearer to users' previous check-ins are ranked higher. However, according to Cheng et al. [2012], the approach for modeling geographical influence may ignore the multimodal nature of users' locations. Therefore, Cheng et al. addressed this problem by explicitly modeling multimodal characteristics. Unlike these existing works, we conduct a two-dimensional kernel density estimation directly, instead of making any assumptions on the spatial distribution. Moreover, based on an influence propagation scheme, we greatly reduce the time complexity of inferring geographical influence. Another difference from these existing approaches is that we apply learning-to-rank techniques to integrate different factors without much human effort for parameter tunings.

Besides the traditional prediction algorithm and the recommendation algorithm, Exploration Prediction is also a core component of the proposed CEPR model. This problem is related to the widely studied novelty and serendipity in the recommendation field [Hurley and Zhang 2011; Vargas and Castells 2011; Ge et al. 2010]. In this work, the designed algorithms attempted to help people discover unusual or unexpected items. Thus, these concepts are different from ours since novel locations in our case are defined as previously unvisited (maybe not unusual) w.r.t a certain user. Moreover, Exploration Prediction forecasts whether a user will visit a novel location or not instead of finding such a novel location. The discovery of novel locations is achieved by the recommendation algorithms. In this way, the recommendation algorithms addressing the novelty or serendipity of items may also serve well for the proposed algorithm. Also, a similar problem to Exploration Prediction has been studied in Song et al. [2010a] and Szell et al. [2012]. In the former one, the probability of exploration was assumed to depend on the number of each user's visited locations, while in the latter one, the exploration probability was simply supposed invariant. Additionally, Exploration Prediction is also related to detecting the deviations from routines [McInerney et al. 2013a], where the author studied how to predict future deviations from routines (including the visit to novel locations and the visit to familiar locations at unusual times) based on time and the immediately preceding state (deviated or not). Our work is different from this existing work in the following ways. First, our work doesn't target the visit to regular locations at unusual times but the visit to novel locations. Second, it assumes that exploration probability is not only different from person to person but also changed over time. Third, it is boiled down to a binary classification problem and thus any features can be easily incorporated. Last but not least, to the best of our knowledge, it is the first time to be applied for connecting location prediction and location recommendation.

3. CHECK-IN DATA ANALYSIS

Since both CEPR and EP are studied on check-ins in this article, we will consider check-ins as location-visiting records and describe the EP problem and the CEPR model using check-ins. Actually, these descriptions can be easily generalized to other types of location-visiting records. Before describing them in detail, we need to perform data analysis on check-ins (1) to understand the existence of regularity and historical dependence in check-in behavior since both of them will be incorporated into our prediction algorithms—however, because the latter one has been carefully studied in Wang and Huberman [2012], it will not be discussed more here; (2) to see whether novelty seeking is ubiquitous in check-in behaviors or not so as to confirm the need for leveraging collaborative filtering for prediction; and (3) to get clarification about some defects of directly applying location recommendation for prediction in order to motivate us to model exploration behaviors and regular behaviors separately. Moreover, since the spatial and temporal characteristics of consecutive check-ins would be useful for the EP problem and recommendation and prediction algorithms, we also perform the spatial-temporal analysis on check-ins. This analysis is conducted on the following two check-in datasets.

One of them is from Jiepang,[1] which is a Chinese location-based social network similar to Foursquare. For the sake of protecting privacy, on both Jiepang and Foursquare, users' historical check-ins are not publicly available. Thus, we cannot directly obtain users' check-ins from these LBSNs. However, users may share their check-ins as tweets on other social networking platforms, such as Weibo and Twitter. For example, Jiepang check-ins are synchronized on Weibo as a particular type of tweet (called location tweets), and Foursquare check-ins are shared on Twitter as normal tweets. Thus, these check-ins can be crawled from these social networking platforms via their open APIs. Some check-in datasets were also crawled in this way [Noulas et al. 2012a; Gao et al. 2012a]. We crawled 36,143,085 Jiepang check-ins at 1,000,457 POIs from 454,375 users via the Weibo API from March 2011 to March 2013, where each user has 80 check-ins and checks in at 47 POIs on average. If we distribute these check-ins into their date, we find that each user only makes 1.5 check-ins each day on average. If we distribute these POIs into 3km^2 regions, each region owns 13 POIs on average and up to 13,068 POIs in the maximal case. For the sake of studying the effect of friendship on location recommendation, we also crawled bilateral friends on Weibo for each user and obtained 3,915,650 social links in total.

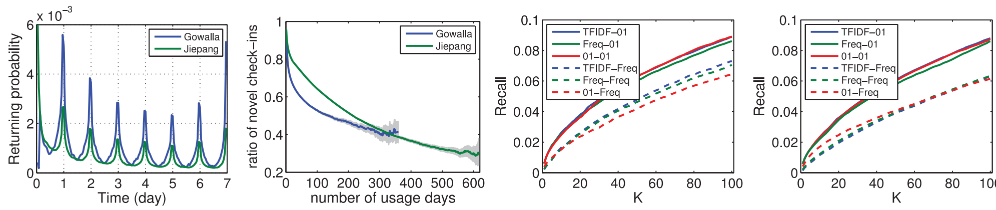
The other one, used in Cho et al. [2011] and crawled from Gowalla[2] from February 2009 to October 2010, contains 6,423,854 check-ins at 1,280,969 POIs from 107,092 users, where each user has 60 check-ins and checks in at 37 POIs on average. If we distribute these check-ins into their date, we find that each user only makes 2.1 check-ins each day on average. If we distribute these POIs into 3km^2 regions, each region owns seven POIs on average and up to 3,940 POIs in the maximal case. This dataset also has 950,327 bidirectional links in companion.

Since EP and the location prediction model of users with a few check-ins are not well trained, we empirically remove users who have fewer than 50 check-ins. As a result, 144,053 users are kept and 887,736 links remain among them on the Jiepang dataset, while on the Gowalla dataset, there are 27,693 users left behind and 111,271 social links remain among them.

Additionally, our subsequent analysis depends on some terms, which need to be defined first. Assume that there are M users in the set U and N locations in the set P. Each check-in c is defined as a triple $\langle u, t, p\rangle$ representing a user $u \in U$ who claims that he or she has visited a location $p \in P$ at time t. We then use $f_{u,p}$ to record check-in frequency and $b_{u,p}$ to memorize the fact that the user u has checked in at the location p. The check-in history of the user u is represented as C_u, and $C_u^t \subset C_u$ is the check-in history before time t, that is, $C_u^t = \{c \in C_u | c.t < t\}$. P_u^t represents the set of locations in C_u^t, that is, $P_u^t = \{p \in P | \exists c \in C_u^t, c.p = p\}$. Based on these notations, we next define novel and regular locations.

Definition 3.1 *(Novel Location).* A location $p \in P$ is *novel* with respect to a user u at time t if it is subject to $p \notin P_u^t$. In other words, *novel* locations with respect to a user at a timestamp are the locations where he or she has never checked in before this timestamp.

[1] www.jiepang.com.
[2] It was acquired by Facebook in 2011.



(a) Regularity (b) Novel Ratio (c) Gowalla-Frequency (d) Jiepang-Frequency

Fig. 1. (a) Returning probability. (b) The ratio of novel check-ins to all check-ins given the number of days of check-in history observed. (c) and (d) Recommendation performance on Gowalla and Jiepang, where each entry denotes the used similarity metric (before "-") and score function: "01," "Freq," and "TFIDF" are binary vectors, frequency vectors, and TF-IDF vectors to represent users, respectively.

Similar to the definition of *novel* locations, we can also define *regular* locations for a user as those that have been checked in before by him or her. Then we can determine the location of each check-in as novel or not. In particular, for any check-in $c \in C_u$ of the user u, its location $c.p$ is *novel* at time $c.t$ if $c.p \notin P_u^{c.t}$ and *regular* otherwise. The corresponding check-ins are also classified into *novel* or *regular*. Given these notations and terms, in the following, we will present data analysis on those two check-in datasets.

3.1. Regularity

By compiling simple statistics on individual check-in frequency at locations, we find that most locations are checked in to only a few times, while several locations are checked in to multiple times. In other words, a majority of check-ins are logged at a small number of locations [Gao et al. 2012a]. In order to investigate the potential regularity underlying the small number of repetitive check-ins, we consider the returning probability [Gonzalez et al. 2008] of each user, defined as the probability that a user will recheck in at a POI t hours after his or her first check-in at the POI. The returning probability on both datasets is plotted in Figure 1(a), where this quantity is characterized by peaks of each day, capturing a strong tendency to recheck in daily at *regular* locations. It thus confirms the existence of regularity so that it will be effective to introduce it into the prediction model. Moreover, although similar trends are manifested on both datasets, there are subtle differences in the strength of regularity, implying the difference of the prediction performance based on regularity (can be validated by our later experiments). In particular, after half a day, the curve of Gowalla is above Jiepang, indicating that the regularity on Gowalla is stronger than on Jiepang so that the prediction performance of regularity on Gowalla should be higher than on Jiepang. However, on Jiepang, a significant peak appears around the first hour, which indicates that users often repeat check-ins at the same POIs within a short time.

3.2. Check-ins at Novel Locations

It is already shown that check-ins exhibit a certain degree of regularity. However, compared to cell tower traces [Gonzalez et al. 2008], the extent of regularity on check-in traces is much smaller. For the sake of studying the potential reasons, we investigate the possibility of checking in at *novel* locations and measure it using a *novel ratio*, which is defined as the ratio of *novel* check-ins to the total number of check-in before a timestamp t. The trend of how the novel ratio varies with t is shown in Figure 1(b). At the beginning of check-in service usage, almost all check-ins are *novel*. As time goes by, the ratio of *novel* check-ins to all check-ins decreases since users will follow their past behavior to check in at *regular* locations, but the declining speed of the novel ratio slows down gradually so that it still remains higher than 40% within 1 year on both datasets. However, the declining speed on Gowalla is much larger than on Jiepang, which means that after a period of time (e.g., 3 months), users on Jiepang are still more likely to try check-ins at novel locations, while users on Gowalla are more inclined to follow their past behaviors. Therefore, comparing these two datasets, it seems that users on Gowalla can be more accurately predicted by their own check-in history (i.e., regularity). This conclusion can be validated in the experimental results presented later. Although there are differences in the declining speed of the novel ratio on two datasets, there is an intersection around the first year of check-in service usage, which indicates that after 1 year of usage, the probability of seeking novel locations on both datasets is close to each other. After this intersection, the novel ratio on Jiepang decreases further but still remains above 25%. Such a significant portion of *novel* check-ins poses great challenges to traditional location prediction since they mainly rely on each user's repetitive check-in behavior.

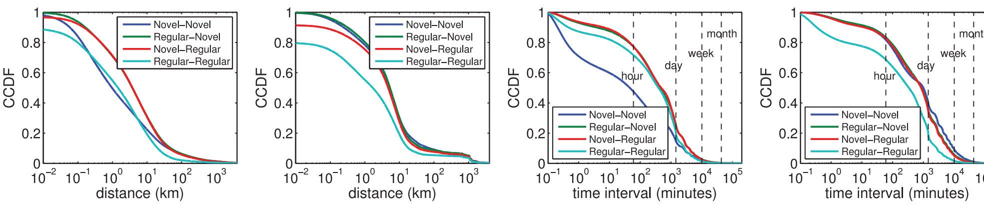
3.3. Effect of Frequency on Location Recommendation and Prediction

To solve the problem stemming from numerous check-ins at novel locations, someone may put forward aggregating all users' history to get a popularity-based model since he or she considers that people usually explore popular locations. However, due to lack of personalization, its contribution in finding the novel candidates is small compared to other factors. Another one may propose to leverage a general collaborative filtering framework (e.g., the sequential collaborative filtering proposed in Lian et al. [2013]) for location prediction, no matter whether the next one is novel or not. However, according to our observation, these solutions don't work well, because we note that most of the existing location recommendation algorithms tend to find unvisited but tailored locations for users and it is difficult for them to capture regularity in human mobility. Inversely, if new recommendation algorithms are designed to model regularity directly, their capacity for finding novel and suitable locations for users becomes weaker, because these models may not well estimate the similarity between users due to the scarcity of mobility data, and the recommending locations for users will be dominated by the regular ones of their similar users. The latter reason will greatly affect location recommendation. For example, assuming that user A and B regularly visit their distinct residences and workplaces and that A has also been to two restaurants x and y occasionally while B has been to restaurants x and z at times, the recommending algorithms capturing regularity tend to recommend the residence and workplace of user B rather than the restaurant z to user A, which may be difficult to accept for user A.

To be more specific, we build a recommendation model, trying to capture regularity. Assume the regularity is characterized by individual check-in frequency $f_{u,i}$ since it has played an important part in traditional prediction algorithms [Gao et al. 2012a; Noulas et al. 2012a; Cheng et al. 2011]. Based on this regularity, the similarity $s_{u,v}$ between user u and v can be easily measured by, for example, cosine similarity. Then, a traditional scoring function $r_{u,i}$ of user u due to collaborative social knowledge is represented as

$$r_{u,i} = \sum_{v \neq u} s_{u,v} f_{v,i}. \tag{1}$$

Actually, such scoring functions for leveraging collaborative social knowledge has also been used in Gao et al. [2012a] and Noulas et al. [2012b]. In order to see its recommendation performance, we split the check-in history into a train (70%) and test part (30%) in chronological order and measure the performance as *Recall*, which indicates what percentage of novel locations in the test part is discovered. For the sake of checking the effect of modeling regularity on recommendation, we compare this model with a



(a) Gowalla-Distance (b) Jiepang-Distance (c) Gowalla-Interval (d) Jiepang-Interval

Fig. 2. (a) and (b) Complementary cumulative distribution function (CCDF) of distance between consecutive check-ins on Gowalla and Jiepang, respectively, where Novel-Regular means that the current location is novel while the next one is regular. (c) and (d) CCDF of corresponding time interval.

similar one in Ye et al. [2011], which ignores users' regularity but only memorizes whether users have been to these locations, that is, using $b_{u,i}$ instead of $f_{u,i}$ in Equation (1). The results of comparison are shown in Figures 1(c) and 1(d), where the former model is indicated by a suffix "freq" while the latter one is indicated by a suffix "01." Moreover, to investigate the effect of regularity modeling on measuring similarity, we compare the aforementioned similarity metric (with prefix "Freq-") with another two measures. One is related to the number of commonly visited locations (with prefix "01-"), and the other is cosine similarity on penalized frequency by inverse user frequency (with prefix "TFIDF-"), similar to cosine similarity between documents based on TF-IDF. From these two figures, it is easily observed that the recommendation model encoding regularity doesn't work as well as the other one, using any similarity measure. Moreover, on the "01" recommendation model, the similarity measure defined based on regularity is also not as good as the other two, though their gap is not large. One possible reason for their small gap is that users check in at a small number of locations so that the density of the user–POI matrix is low [Ye et al. 2011]. Thus, based on these two observations, we see that the recommendation algorithm modeling regularity directly cannot achieve the best performance of recommendation. To this end, we build the location recommendation and prediction model separately and integrate them via Exploration Prediction in the Collaborative Exploration and Periodically Returning model.

3.4. Spatiotemporal Characteristics

After motivating the need to model novel and regular location prediction separately, we further analyze the characteristics of check-ins in this situation from the spatiotemporal perspectives, with the aim of understanding their effect on location recommendation and prediction as well as Exploration Prediction.

From the spatial perspective, since the distribution of distance between consecutive check-ins has been studied for filtering distant locations for recommendation and prediction [Noulas et al. 2012a; Cho et al. 2011; Ye et al. 2011], we are also interested in them, but we will distinguish novel locations from regular ones. Thus, we plot the complementary cumulative distribution function (CCDF) of this distribution in Figures 2(a) and 2(b). From them, several interesting observations can be made. First, most check-ins (over 80%) are within 10 kilometers from the immediately preceding locations. Hence, when recommending or predicting the next location, the candidate locations can be restricted to within 10 kilometers from previous ones. Second, when we already know that users have checked in at regular locations, the next regular locations are nearer to them than the next novel locations on both datasets. Based on a hypothesis testing–unpaired two-sample t-test, such an observation is significant. It not only shows that in order to explore nearby locations users are willing to take the cost of distance and visit farther locations but also indicates that users frequently

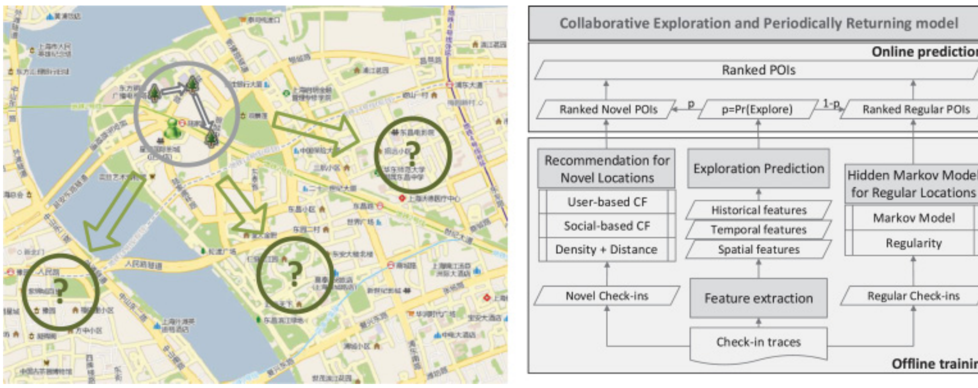


Fig. 3. (a) A typical scenario for next check-in location prediction. After three successive check-ins (tree icons) of a user (head icon), we predict his or her next check-in location. (b) The framework for exploiting Exploration Prediction to combine location prediction and location recommendation.

explore the neighborhood of their familiar locations. Finally, when previous locations are novel, CCDF exhibits distinct trends on both datasets. On the Gowalla dataset, users are more willing to explore continuously. In other words, the distance between consecutive novel check-ins is smaller than the distance from novel check-ins to regular locations, and this relationship is significant according to an unpaired two-sample t-test. This means that when a user has visited a new attraction, he or she may also try a nearby restaurant. However, on the Jiepang dataset, we don't observe such a trend. Based on this analysis, we can see that the distance between consecutive locations has a different effect on location recommendation and prediction, and that their distribution in these four scenarios has some different points. It also confirms the need to separate novel locations from regular ones. We will further study its effect on recommendation and prediction in later experiments.

From the temporal perspective, similarly, we study the distribution of time interval between consecutive check-ins to see whether there are discrepancies when differentiating novel locations from regular ones. If this is true, its distribution can be used for Exploration Prediction. Its distribution in terms of CCDF is shown in Figures 2(c) and 2(d). From them, we can make the following two observations. First, given that current locations are regular, their time interval to the next regular locations is smaller than the next novel ones, although the gap between them on Jiepang is larger than on Gowalla. When performing an unpaired two-sample t-test, such a relationship is significant. This implies that when a user has visited a regular location, if the time interval is small, this user is less inclined to explore. Second, if current locations are novel, their time interval to the next novel locations is significantly smaller than the next regular ones on Gowalla, while on Jiepang, this difference is not significant according to an unpaired two-sample t-test. It implies that on Gowalla, users will be more likely to check in at novel neighboring locations consecutively in a short time (e.g., hour), in line with the preceding spatial analysis on Gowalla. Therefore, based on this analysis, we can see the benefit of this distribution to Exploration Prediction.

4. COLLABORATIVE EXPLORATION AND PERIODICALLY RETURNING MODEL FOR LOCATION PREDICTION

4.1. Overview

Our final goal is to predict the next check-in locations for each user, whose typical scenario is shown in Figure 3(a), where the next location of a user (head icon) is going to be predicted after he or she has made three successive check-ins (tree icons). In this task, according to our previous analysis, we need to separate the prediction of novel locations from regular locations. In order to jointly make the best use of their individual power, we put forward a CEPR model to alternate between location recommendation and prediction based on an EP problem, illustrated in Figure 3(b). To be more specific, when people are predicted to do exploration ($Pr(Explore) = 1$), a recommendation algorithm, armed with collaborative social knowledge and assisted by geographical influence, will be applied for seeking suitable candidates; otherwise, ($Pr(Explore) = 0$), a traditional prediction algorithm, which incorporates both regularity and a Markov model, will be put into use for figuring out the most possible locations to visit. When EP outputs an exploration tendency ($Pr(Explore) \in [0, 1]$) instead of fully accurate prediction results ($Pr(Explore) \in \{0, 1\}$), the outputs can be *interpolated* together.

In the following, we first present Exploration Prediction in Section 4.2. We then introduce regularity and a Markov model for regular location prediction in Section 4.3 and build the recommendation model to find novel candidates for each user in Section 4.4. Finally, we propose the Collaborative Exploration and Periodically Returning model based on Exploration Prediction to integrate location recommendation with location prediction in Section 4.5.

4.2. Exploration Prediction

Based on the definition of novel and regular locations, we can formally define Exploration Prediction as follows:

Definition 4.1 (*Exploration Prediction*). Given check-in history C_u of a user u, an Exploration Prediction problem with respect to the user u predicts whether his or her next check-in location is *novel* or not.

Therefore, when predicting the novelty of the next check-in location, we don't need to know the next check-in location in advance but just determine the next check-in location as *novel* or *regular*. Thus, it is boiled down to a binary classification problem, which takes features as inputs and outputs a classification result (novel or not) or exploration tendency (e.g., a probability of classifying the next location as novel). In the classifiers, we consider three types of features, historical, temporal, and spatial features.

First of all, *historical features* not only summarize the personality trait of novelty seeking (i.e, how much users love check-in) but also reflect users' current status of neophilia, including whether they are doing exploration now and how many opportunities are left in which they can seek novel locations. In the latter case of current neophilia status, assuming that a user has visited lots of locations near his or her activity areas, the propensity of seeking novel locations should be smaller. Second, since time information is often assumed available in location prediction [Gao et al. 2012c; Cho et al. 2011; Noulas et al. 2012a; McInerney et al. 2013b] and users usually have distinct degrees of novelty-seeking tendency at different times, we introduce *temporal features* to consider the effect of this temporal information on Exploration Prediction. For example, users may prefer to do exploration during weekends. Moreover, according to the previous analysis, the time interval from the current check-in to the next one also affects Exploration Prediction and thus is also placed into temporal features. Last but not least, we take spatial features into account for Exploration Prediction because users also exhibit different propensities of novelty seeking at locations with distinct degrees of familiarity. For example, if a user has arrived in an unfamiliar location (e.g., city), his or her propensity for novelty seeking will increase. However, spatial features are usually unavailable in the prediction scenario, but they can be used in the case of location naming [Lian and Xie 2011; Shaw et al. 2013] or sensor-augmented

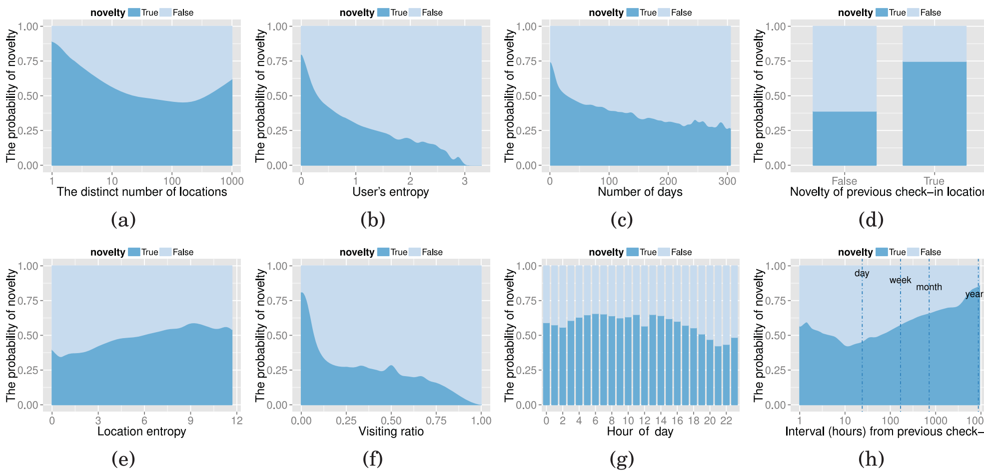


Fig. 4. The effect of different features on the possibility of exploration of novel locations. All x-axes of these figures represent corresponding features, while y-axes are the conditional probability of novelty of next check-in location given the corresponding feature.

mobile phone localization [Azizyan et al. 2009; Ofstad et al. 2008; Zhang et al. 2012]. Therefore, we also assume the knowledge of physical locations available.

Next, we elaborate the feature extraction procedure and study the effect of these features on Exploration Prediction on the Gowalla dataset. This study is achieved by plotting the conditional probability of exploration, that is, checking in at novel locations next, on features and observing how this probability changes with the corresponding features. Without loss of generality, the features in this study are extracted with respect to the check-in $\langle u, t, p \rangle$.

4.2.1. Historical Features. Historical features are extracted from a user's past check-in history, that is, C_u^t. Particularly, we consider the following five features.

(1) The distinct number of locations $S = |P(C_u^t)| \in \mathbb{N}$, which was also used in Song et al. [2010a]. The probability of checking in at novel locations conditioning on S is shown in Figure 4(a). With the increasing number of visited locations, this probability first decreases and the declining speed slows down gradually, similar to the trend in Song et al. [2010a]. The difference lies in that the decline is followed by an increasing trend when the number of locations is over 300. The reason is that there are only a small number (6%) of users checking in at over 300 locations; these users like to check in more than others. Such differences may root from users being able to check in at any location nearby. In spite of this difference, it still indicates that after a user has visited lots of locations near his or her activity areas, his or her propensity of seeking novel locations should be smaller. Thus, this quantity reflects a user's current status of neophilia.

(2) User's entropy $H \in [0, \log_2 S]$, which is calculated based on the user's check-in frequency at POIs in P_u^t. Specifically, if the check-in frequency of the user u at these S POIs is $f_1, \ldots, f_S$, respectively, his or her entropy is $H = -\sum_i \frac{f_i}{\sum_j f_j} \log_2 \frac{f_i}{\sum_j f_j}$. However, a user's entropy is actually correlated to the quantity S. For the sake of studying the heterogeneity of check-in frequency distribution among different users, we eliminate the effect of the distinct number of locations. In other words, we consider $\log_2 S - H$ as the feature of the user's entropy. We show its effect on Exploration Prediction in Figure 4(b). When this quantity is smaller, indicating that the user's check-in

frequency is more uniformly distributed on POIs, the probability of checking in at novel locations is higher. The underlying reason is that more uniform distribution over POIs indicates that their average check-in frequency is close to 1 since users' average check-in frequency on POIs is less than $1.62(\frac{60}{37})$. In other words, users with higher entropy prefer to check in at novel locations. Due to this property, it reflects a user's personality trait of novelty seeking.

(3) Novel ratio $Ratio \in [0, 1]$, which is the ratio of the number of check-ins at novel locations to the total number of check-ins with respect to user u before time t and thus indicates a user's current status of novelty seeking. Its conditional probability plot is not shown here since the observations we can make are in line with our assumption. That is, the larger the novel ratio is, the higher the probability is that the users will check in at novel locations.

(4) Number of days $\#Days \in \mathbb{N}$ in a user's check-in history C_u^t. Its impact on Exploration Prediction is illustrated in Figure 4(c). We observe that as long as users use check-in services for a longer period of time, they will check in at novel locations with lower probabilities. Thus, its effect is similar to the distinct number of locations and thus it still reflects a user's current status of novelty seeking. However, by comparing their variance among the population, the distinct number of locations is varied much more from person to person than is the number of days. In this sense, it seems more appropriate to consider the number of days instead of the distinct number of locations as the independent variable in the exploration model [Song et al. 2010a] on the check-in datasets. In other words, the probability of doing location exploration is in proportion to $t^{-\gamma}$ instead of $S^{-\gamma}$, where t is the number of days for which a person has used the check-in services.

(5) Novelty of previous check-in location, $N_{t-1} \in \{False, True\}$. Its effect on the novelty of the next location is shown in Figure 4(d). When the location of the previous check-in is *novel*, the user will check in at a novel location next with a significantly higher probability; when the location of the previous check-in is *regular*, users tend to check in at *regular* POIs next. This is in line with the spatial-temporal analysis that on Gowalla users tend to explore continuously and visit the neighborhood of their familiar locations together.

4.2.2. Spatial Features. Spatial features are based on physical locations and include the following three features.

(1) The average distance d_l from the user's previous check-ins. We omit its conditional probability plot due to its small effect, particularly when the distance is within 10km. However, when the distance increases to more than 10km, the probability that users check in at novel locations next becomes higher. This is because when users are very distant from their previous locations, it seems that they have already arrived at a location with which they may not be familiar.

(2) Location entropy H_l. Given a physical location l, we first calculate each user's check-in frequency within 3km from l (here 3km represents the neighborhood of a location; other options can also be set as long as the neighborhood includes sufficient visiting users and the area of the neighborhood is not large). If we assume that M_l users have checked in around this location and their check-in frequency is $u_1, \ldots, u_{M_l}$, then the entropy of this location is $H_l = -\sum_i \frac{u_i}{\sum_j u_j} \log_2 \frac{u_i}{\sum_j u_j}$. For locations with higher entropy, it is not only likely that users' check-in frequency at these places is more uniformly distributed but also possible that there are a larger number of visiting users at these places. However, in the check-in data, around the locations of higher entropy, the check-in frequency of visiting users is usually small so that a large number of visiting users only make a small number (close to 1) of check-ins. Such locations can be attractions and public transportation, for example, airports or railway stations. Around

these locations, it is more likely that users will check in at novel locations according to Figure 4(e), though its effect on the novelty of the next location is not large. In contrast, at the locations of lower entropy, since most check-ins around are usually made by a smaller number of users, it is more likely that these users will check in at these locations again, as shown in this figure.

(3) Visiting ratio $Ratio_l$, which is the ratio of the number of check-in locations to the total number of locations within 3km around physical location l. This quantity describes how many POIs nearby have been checked in to before. If it is higher, most locations have been checked in to before so that the user u may not try to check in at the remaining novel locations. If it is lower, there are still a larger number of unchecked-in locations nearby, and thus there is a higher probability that the user will seek novel locations. These trends are shown in Figure 4(f). Moreover, the probability of checking in at *novel* locations reaches a comparatively stable state around $Ratio_l = 0.125$. In other words, after users search around and check in at several interesting locations, they do not tend to continue exploration since it's difficult for them to discover interesting novel locations.

4.2.3. Temporal Features. Temporal features are extracted according to the temporal information and include the following four features.

(1) Hour of day, $HOD \in \{0, \ldots, 23\}$. At different hours of the day, the probability of checking in at novel locations is different. Thus, we plot the conditional probability with respect to hour of day in Figure 4(g). It indicates that at noon (having lunch) and in the evening (at home), users tend to be more regular.

(2) Day of week, $DOW \in \{0, \ldots, 6\}$. Similar to hour of day, the probability of checking in at novel locations is also varied during different days of the week. This result is not shown because it can be consistent with our assumption in practice. That is, users are more likely to visit and check in at novel locations during weekends since they have more time to choose leisurely locations.

(3) Hour of week, $HOW \in \{0, \ldots, 167\}$. For the sake of seeing the effect of taking hour of day and day of week as a whole, we combine them together as a single variable, which can be considered as the two-order interaction in the analysis of variance. We don't show its effect since its general trend is similar to case (1) and case (2).

(4) The time interval in hours from a previous check-in. Its effect on the probability of checking in at novel locations is plotted in Figure 4(h). When the time interval is in 1 day, the probability of checking in at regular locations is higher and it peaks around 12 hours. It is compatible with the daily regularity observed in check-in datasets [Cheng et al. 2011].

With these features at hand, the next step is to train a classifier to map the extracted features of each check-in to the novelty of its location. We consider two supervised models to achieve this goal: logistic regression (LR) and Classification and Regression Tree (CART). Before applying LR, since not all the introduced features are in the same scale, we make a transformation using the logarithm function if any. The reason for using these two models is that we would like to know whether nonlinear classifiers could be better than linear ones for Exploration Prediction. Actually, we can also leverage any other pairs of linear and nonlinear supervised models. Thus, the training time complexity of Exploration Prediction depends on the used classifier, being usually completed offline. As for the output of classifiers, according to the need in the Collaborative Exploration and Periodically Returning model, we consider two cases. One of them is the result of classification, which determines whether a user will check in at novel locations next. The other one is a probability value, determining the tendency of checking in at novel locations.

4.3. Hidden Markov Model for Regular Location Prediction

For the prediction of regular locations, we need to fully exploit the time-dependent regularity and a Markov model according to the previous analysis in Section 3. For the sake of being estimated well, a Markov model is assumed to be the first order. These two components can occur simultaneously if we consider the location as hidden states and the temporal information as the observations of a hidden Markov model. Thus, our solution can be considered as the integration between regularity [Song et al. 2010b] and a Markov model [Song et al. 2004]. As for the temporal information, we only use hour of day and day of week as categorical variables and assume the independence between them given the location (hidden variable). Since hidden states (locations) in the check-in traces are already known, we leverage supervised learning to estimate the parameters. Due to the assumption in HMM, the estimation of these three parameters can be achieved separately. When estimating these parameters, we are trying to address the overfitting problem due to the scarcity of individual check-ins and the large number of visited locations.

The first probability to be estimated is the emission probability, that is, $P(h|l)$ and $P(d|l)$, where d, h, and l are the d^{th} day of the week, the h^{th} hour of the day, and the location l, respectively. These parameters are first initialized by a maximum likelihood estimation as follows: $P_{ML}(h|l) = \frac{n(h,l)}{\sum_h n(h,l)}$ and $P_{ML}(d|l) = \frac{n(d,l)}{\sum_d n(d,l)}$, where $n(h,l)$ is the number of check-ins at location l at the h^{th} hour of the day, and $n(d,l)$ is the number of check-ins at location l at the d^{th} day of the week. Without sufficient training data, such estimation of these probabilities cannot guarantee that the difference of the probability between neighbor hours of the day and the difference of the probability between neighbor days of the week should be small. For example, we observe that a user checked in at a Chinese restaurant at 6 p.m. only once. If this user would return to this restaurant in the near future, the probability of its check-in over hours of day should be distributed around 6 p.m. (e.g., Gaussian centering at 6 p.m.) rather than at 6 p.m. exactly. Thus, these parameters are further transformed by a Gaussian kernel smoothing function as follows:

$$
\begin{aligned}
\tilde{P}(h|l) &= \frac{\sum_{g=0}^{23} K(\frac{d(h,g)}{\sigma_{g,l}}) P_{ML}(g|l)}{\sum_{h'=0}^{23}\sum_{g=0}^{23} K(\frac{d(h',g)}{\sigma_{g,l}}) P_{ML}(g|l)} \\
\tilde{P}(d|l) &= \frac{\sum_{e=0}^{6} K(\frac{d(d,e)}{\sigma_{e,l}}) P_{ML}(e|l)}{\sum_{d'=0}^{6}\sum_{e=0}^{6} K(\frac{d(d',e)}{\sigma_{e,l}}) P_{ML}(e|l)},
\end{aligned}
\tag{2}
$$

where $d(h,g) = \min(|h-g|, 24-|h-g|)$ is the distance between the h^{th} and g^{th} hour of the day and $d(d,e) = \min(|d-e|, 7-|d-e|)$ is the distance between the d^{th} and e^{th} day of the week. The reason for defining distance in this way is that there is a cyclic property among them (the probability of 0 a.m. is close to 1 a.m. and 23 p.m. and the probability of Sunday is also close to Saturday and Monday). $K(x)$ is a truncated standard Gaussian distribution over $x \in [0, +\infty)$. $\sigma_{g,l} = \sigma n(g,l)^{-\frac{1}{5}}$ is a kernel radius similar to the bandwidth in KDE. The form we choose follows the idea of practical selection of the bandwidth ($1.06\hat{\sigma} n^{-\frac{1}{5}}$) in KDE [Dehnad 1987]. We assume that σ is unknown but it will be tuned on the validation set instead of being estimated as MLE $\hat{\sigma}$ since sometimes $\hat{\sigma}$ cannot be estimated (e.g., location checked in only once). The time complexity for such an estimation is $O(N_u)$, in linear proportion to the number of visited locations N_u.

The second parameter to be estimated is the initial probability of states, which is simply set as maximum log-likelihood estimation, that is, $p_{ML}(l) = \frac{n(l)}{\sum_l n(l)}$, where $n(l)$ is the check-in frequency of location l. Its estimating complexity is also $O(N_u)$.

Finally, we estimate the transition probability across hidden states. However, if it is directly estimated via maximizing the log-likelihood, it suffers from the overfitting problem due to the insufficiency of training data. According to the two aforementioned check-in datasets, the number of parameters is around 40 × 40 (there are 40 POIs for each user on average), while there are only about 60 training instances (check-ins) for each user. It can be smoothed by the Laplace smoothing techniques, assigning a nonzero probability to events that are not observed till now but that may happen in the future, whereas such smoothing techniques don't differentiate the events of the same observed frequency. For example, if a user u visits both a location B and a location C three times after visiting a location A, the probability of visiting B and C after A is the same. Nevertheless, if the user u visits B more often than C, the probability of visiting B conditioning on visiting A should be higher. To this end, we leverage the widely used Kneser-Ney smoothing techniques for language modeling [Chen and Goodman 1996]. In particular, $P(l|k)$ is derived as

$$P(l|k) = \frac{\max\{n(k,l)-\delta, 0\}}{\sum_{l'} n(k,l')} + \frac{\delta \sum_{l'} \mathbf{1}_{\{n(k,l')>0\}}}{\sum_{l'} n(k,l')} \frac{\sum_{p} \mathbf{1}_{\{n(p,l)>0\}}}{\sum_{l'}\sum_{p} \mathbf{1}_{\{n(p,l')>0\}}}, \tag{3}$$

where $\mathbf{1}_{\{\cdot\}}$ is an indication function and $0 \le \delta \le 1$ is a discounting parameter that can be set using the empirical formula $\delta = \frac{n_1}{n_1+2n_2}$ (n_1 and n_2 are the number of one-time transitions and two-times transitions across locations, respectively). The basic intuition of this equation is to discount the observed times of transition and turn them over to the possibility that some locations cannot be transited given location k. Moreover, this estimation of transition probability ensures zero-order distribution such that the marginal of the first-order probability distribution matches the marginals of the training data. Specifically,

$$\sum_{k} P(l|k) P_{ML}(k) = P_{ML}(l). \tag{4}$$

Therefore, $P_{ML}(l)$ is the stationary distribution of a Markov process determined by stochastic transition matrix $P(l|k)$. For more information and a detailed deviation of these equations, please refer to Chen and Goodman [1996] and Teh [2006a]. The complexity for such an estimate is $O(N_u^2)$.

4.4. Recommendation for Novel Location Prediction

If the next location is predicted as novel, we aim to discover the potentially interesting and tailored locations for users. Here, we exploit three features in the state-of-the-art location recommendation for this purpose, similar to that in Ye et al. [2011], Shani et al. [2005], Noulas et al. [2012b], and Cheng et al. [2012]. Two of them leverage collaborative social knowledge by means of collaborative filtering. The first one is user-based collaborative filtering based on the similarity between users on location visit history. According to the previous study, a user checking in at locations contributes more to location recommendation than its frequency. Therefore, we represent a user u as $\mathbf{b}_u \in \{0, 1\}^N$, whose each entry indicates whether he or she has been to the corresponding location. We then compute the similarity between user u and v as $s^l_{u,v} = \mathbf{b}_u^T \mathbf{b}_v / \|\mathbf{b}_u\| \|\mathbf{b}_v\|$. Here, the overall complexity for computing similarity between users is $O(M\|\mathbf{B}\|_0)$, where the l_0 norm of the matrix equals the number of its nonzero entries and $\mathbf{B} \in \{0, 1\}^{M\times N}$ is a user–POI 0/1 rating matrix obtained by first stacking $b_1, \ldots, b_M$ by column and then transposing. For more detailed analysis, please refer to Desrosiers and Karypis [2011]. After that, the scoring function via this similarity metric is $\mathbf{s}_u^T \mathbf{B}$, which costs $O(\frac{\|\mathbf{s}_u\|_0}{M}\|\mathbf{B}\|_0)$ and thus depends on the number of similar users of the use u. The second feature is based on the similarity between users on a social network. The similarity between users is defined as $s^f_{u,v} = |\{\text{common friends between u and v}\}|/|\{\text{all friends of both u and v}\}|$. We still use the rating matrix $\mathbf{B}$ in scoring function together with $s^f_{u,v}$ to get a final score function. Its complexity analysis is similar to the previous feature. However, due to the sparsity of social network information, the complexity of social-based CF is usually smaller than user-based CF.

In addition to those two features related to collaborative filtering, we also exploit geographical features for location recommendation. They include not only the geographical influence, related to the spatial clustering phenomenon in users' check-in behavior, in existing algorithms but also an unusual feature in recommendation, the distance from immediately preceding locations. The reason for the unusualness of the latter feature in recommendation is that existing work is usually evaluated with ignoring chronological order of check-ins so that this distance is not well defined. In addition to providing distance, the existence of previous locations can also supply other information, such as the activity that users conduct just now. Therefore, we can also leverage this additional information for location recommendation, similar to that in Lian et al. [2013], but due to it being beyond the focus of this article, we will not discuss it here.

To infer the geographical influence, we don't follow the existing work to assume the power law [Ye et al. 2011] of distance between any pair of locations or Gaussian Mixture distribution [Cheng et al. 2012] over locations. Instead, we directly perform two-dimensional *kernel density estimation* for individual *spatial distribution*, since it can also agree with the observation that the POIs visited by the same user tend to be clustered geographically. More importantly, its complexity can be greatly reduced by means of an influence propagation scheme. To the best of our knowledge, this is the first attempt to infer geographical influence by means of two-dimensional kernel density estimation. In the following, we elaborate this procedure.

Following the definition of kernel density estimation, the density at location l with respect to a user u, who has checked in at N_u POIs, is represented as

$$p_u(l) = \frac{1}{N_u h} \sum_{n=1}^{N_u} K\left(\frac{d(l_n, l)}{h}\right), \tag{5}$$

where $K(\cdot)$ is a kernel function and $d(l_n, l)$ is the distance function between location l_n and l. Thus, it costs $O((N - N_u)N_u)$ to get the geographical influence of all the remaining $N - N_u$ POIs. This is almost the same as the cost in power law distribution of distance between any pair of POIs, except that it doesn't have $O(N_u^2)$ additional cost to compute distance between pairs of observed locations. In addition to the difference in time complexity, the power law assumption slows down the geographical influence with the increase of distance compared to a normal kernel, which is a widely used kernel function. It indicates that farther locations (e.g., 10km) from existing ones can still have a significant influence under the power law assumption. However, if we can leverage the kernel function with a long tail, such as t-distribution, similar cases can also be captured in kernel density estimation.

Since the time complexity of the current approach is $O(N_u(N - N_u))$, it can be reduced if we approximate kernel density estimation by only taking into account candidate locations within d km (set as 10km according to Figures 2(a) and 2(b)) from any existing one. The setting of bandwidth h requires that the influence of candidate locations on the border is close to zero, for example, $K(\frac{d}{h}) < \epsilon K(0)$ ($\epsilon \ll 1$). In this case, to estimate the density of location l, only the observed locations within d km from l are taken into account. However, when no observed location is within d km from l, the range query on l still needs to be performed. To avoid such unnecessary computation, we propose a propagation scheme to infer the geographical influence for candidate locations. In particular, for each observed location l_n, its geographical influence is propagated to all candidate locations within d km. In particular, each candidate l receives $\frac{1}{h}K(\frac{d(l_n,l)}{h})$ influence from l_n. Then, by aggregating the received influence and dividing it by N_u, the density of each candidate location can be obtained. In regard to its time complexity, it consists of two parts. The first part is to retrieve POIs within a square of length d km given N_u location queries, the time complexity of which depends on the usage of the spatial index. If we perform these range queries with the help of a range tree [Bentley and Maurer 1980], its time complexity is $O(N_u(\log^2 N + K))$, where K is the number of retrieved locations. The second part is computing the geographical influence for these K retrieved POIs, the time complexity of which is $O(N_u K)$. Therefore, propagation-based kernel density estimation can greatly reduce the time complexity for inferring geographical influence compared to the power law solution. Actually, this solution still includes some redundant computation. For example, the geographical influence of two neighbor POIs should be close to each other, but their influences are actually calculated twice in total. These redundancies can be alleviated by first splitting the whole world into many square grids of length r (e.g., $r = 0.1km$) and supposing that POIs within the same grid have the same influence. As long as r is small enough, the performance of inferring geographical influence is not affected. Due to this being beyond the scope of this article, we don't discuss it in detail.

After obtaining these features, in order to jointly leverage them to improve the recommendation performance, we put them into a supervised learning framework. However, we are only aware of where users have been (positive class) but don't know the locations they don't like or want to go (negative class). This is similar to the one-class problem in recommendation or classification [Pan et al. 2008]. One popular solution for this problem [Rendle et al. 2009] is to randomly sample unvisited locations for each user as negative examples and learn pairwise preference based on these features. Given negative samples, this solution can be in analogy to a learning-to-rank algorithm, that is, RankNet [Burges et al. 2005]. Meanwhile, RankNet has exploited Neural Network to map these features to the pairwise preference in a nonlinear way. Therefore, we adopt RankNet to rank POIs for each user given these features. However, RankNet usually provides a score for each POI, which doesn't necessarily have a probabilistic interpretation. In this case, we transfer them into a probability value using the softmax function if any. Finally, when predicting novel locations, the probability that regular locations of a user appear in his her candidate list should be close to zero. Hence, we remove locations already in the training portion from the list of candidate locations.

4.5. Collaborative Exploration and Periodically Returning Model

Providing the probabilistic output of prediction algorithm $P_r(l)$ (the subscript r indicates regular) and recommendation algorithm $P_n(l)$ (the subscript n indicates novel), we propose the Collaborative Exploration and Periodically Returning model to combine them based on Exploration Prediction $Pr(Explore)$ as follows (here, for simplifying notation, we omit the corresponding context information in algorithms):

$$P(l) = Pr(Explore)P_n(l) + (1 - Pr(Explore))P_r(l). \tag{6}$$

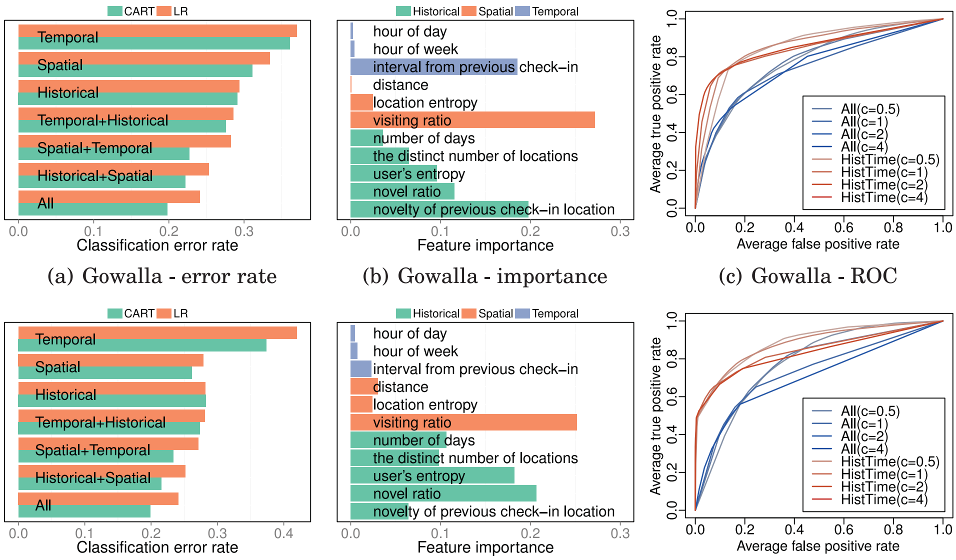
If $Pr(Explore) \in \{0, 1\}$ (i.e., EP predicts the next location as novel or not), we can switch between the location recommendation model and the location prediction model based on Exploration Prediction. In particular, when people are predicted to do exploration $Pr(Explore) = 1$, $P_n(l)$ will be applied for finding novel candidates; otherwise ($Pr(Explore) = 0$), $P_r(l)$ will be put into use for figuring out the most possible locations to visit. Due to the discrete value of $Pr(Explore)$, we denote this case as "hard" integration. If $Pr(Explore) \in [0, 1]$ (i.e., EP outputs the probability of check-in at novel location (Exploration)), we can interpolate the location prediction model with the location recommendation model. In other words, both novel and regular locations are ranked together in this case for the final location prediction. Due to the continuous value of $Pr(Explore)$, we denote this case as “soft” integration. Such a naming rule is analogous to two different ways of clustering: K-Means and Gaussian Mixture Model. The time complexity of obtaining $Pr(Explore)$ also depends on the used classifier, which may be in constant time, and the procedure of extracting features from each user’s past check-ins. Most features can be extracted incrementally or instantly except the spatial features. The extraction of spatial features can be greatly accelerated using the spatial index and thus costs $O(\log_2 N + K + N_u)$, according to the analysis in Section 4.4.

5. EXPERIMENTS

We evaluate the proposed algorithms on two aforementioned check-in datasets. The check-in history of each user is split into a training portion (70%) and a testing portion (30%) in chronological order. We train our models on all users’ training portions and test the performance of our models on all users’ testing portions. Besides, a similar split strategy is applied on the training portion for obtaining a validation dataset and a new training dataset so that some parameters can be tuned on the validation dataset. For assessing the performance of regular location prediction, we measure Accuracy, that is, what percentage of testing check-in locations can be returned with the highest probability. Based on the Accuracy, we compare the proposed algorithm with regularity with different time granularity, including hour of day (*HOD*) and hour of week (*HOW*) [Song et al. 2010b], and a Markov model with different smoothing strategies, including Markov with fallback (*Markov-F*) [Song et al. 2004] and a Bayesian Markov model (*HPYP*) [Gao et al. 2012a]. In HPYP, the discount of HPYP is initialized $d[0, 1] = \{0.8, 0.9\}$ and updated using the Metropolis-Hastings sampling algorithm [Teh 2006a]. The initial setting of $d[1]$ is important since its role is similar to δ in Markov-S, so we try different values and select the best one. For evaluating the performance of novel location prediction, we compute Recall@k, which is widely used in recommendation [Ye et al. 2011; Noulas et al. 2012b] and indicates what percentage of novel locations can be returned at the top k positions. According to Recall, we study the performance of each feature and their combination. Finally, we use Accuracy@k, which calculates what percentage of check-in locations can be returned at the top k positions, to assess the performance of CEPR. Accuracy@k differs from Recall@k, since Recall targets novel check-ins while Accuracy concentrates on both regular and novel check-ins. Based on the Accuracy@k, we address the superiority of CEPR.

5.1. Exploration Prediction

Before evaluating location prediction, we first study the performance of Exploration Prediction. The classification models, including CART and LR, are trained on all users’ training portion of check-ins. The performance of these two algorithms with different configurations of three types of features is shown in Figures 5(a) and 5(d). From these two figures, we can make the following observations: (1) all three types of features are effective for Exploration Prediction since the error rate of a MostFrequent classifier, determining check-in locations as the most frequent label (novel or regular), is 0.42 and 0.49 on Gowalla and Jiepang, respectively, and higher than all three types of features; (2) when comparing three individual types of features, the historical and spatial features perform better than temporal features, implying that the probability of seeking novel locations does not change a lot with time; (3) these three types of features complement each other since every paired combination outperforms the individual’s and a combination of three types of features outperforms any other configuration; and 4) the performance of CART is higher than LR under all configurations of three types of



(a) Gowalla - error rate (b) Gowalla - importance (c) Gowalla - ROC

(d) Jiepang - error rate (e) Jiepang - importance (f) Jiepang - ROC

Fig. 5. The evaluation of Exploration Prediction. In (b) and (e), the relative importance of features is referred to in Equation (10.42) in Hastie et al. [2001]. In (c) and (f), novel location is considered as the positive class and "HistTime" ignores the effect of spatial features, while "ALL" considers all three types of features. Moreover, c is the ratio of the cost of classifying regular locations as novel to the cost of classifying novel ones as regular.

features. The reason may lie in the linear assumption in LR, indicating the nonlinear classifier could be more appropriate for Exploration Prediction. Therefore, it will be used for our later experiments.

After CART has been trained on training data, the relative importance of features (Equation (10.42) in Hastie et al. [2001], summarizing their contribution of constructing this tree) can be obtained and plotted in Figures 5(b) and 5(e). From these two figures, the importance of features on two check-in datasets shows some differences. First, novelty of previous locations and time interval from previous check-ins play much more important parts in the model trained on the Gowalla check-ins. This is consistent with the observation in Figures 2(a) and 2(c) that on Gowalla, users are more willing to explore continuously. Second, a user's entropy and novel ratio are of great importance on Jiepang. This may result from the diverse personality trait of novelty seeking in human mobility on LBSNs among the population. In spite of these differences in the importance of these features, there also exists some features showing similar importance on both datasets. For example, the visiting ratio is significantly important on both datasets. Therefore, if we can know the physical location of the next check-in in advance, we can greatly improve Exploration Prediction. For another example, the effect of categorical temporal information is small compared to the time interval, though hour of week is more important than day of week and hour of day. The reason lies in the following two aspects. First, we aggregate all users' check-in data to train a universal model so that it hides their effects at the individual level. Second, the effect of this categorical temporal information on the probability of novelty is surely not as large as others, as shown in Figure 4. In addition to two previous features with similar importance, we observe that the effect of distance from previous check-ins is small on both datasets. This could be because most of the distances are smaller than 10km according to our observations in Figures 2(a) and 2(b) as well as the results of Noulas et al. [2012a] and

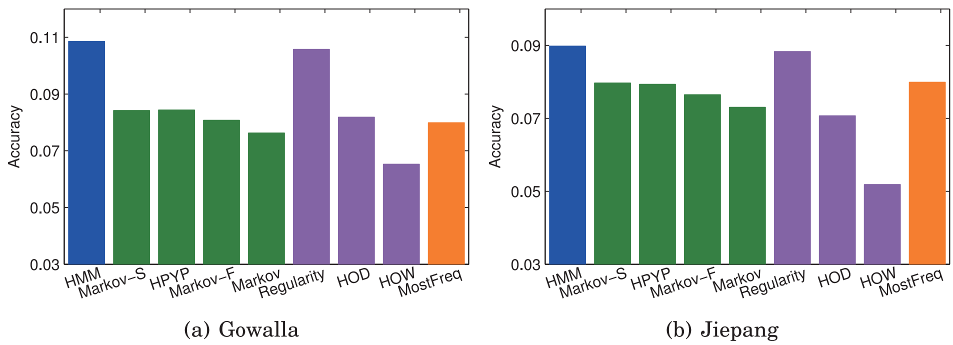


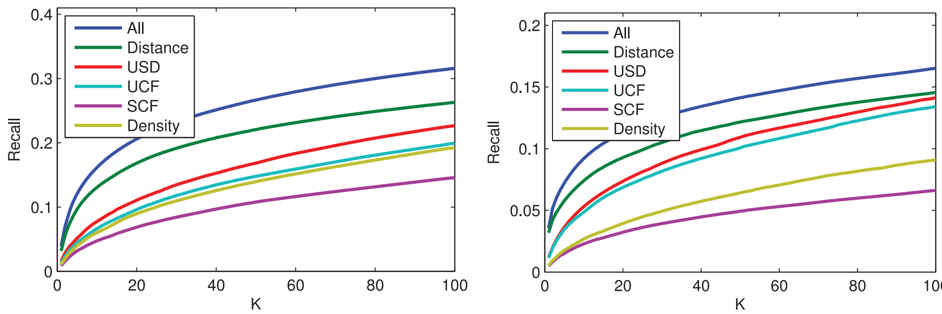
Fig. 6. Regular location prediction performance. Markov-S and Regularity are our smoothed version of the Markov model and time-dependent model for regular location prediction, respectively. And in HMM, we have leveraged the power of both Markov-S and Regularity.

Cheng et al. [2011], while the probability of novelty is almost invariant within this range. Finally, both the number of days and the distinct number of locations don't play a very important role on both datasets. Thus, using the distinct number of locations or the number of days [Song et al. 2010a] to determine the probability of exploration is not enough.

Next we investigate the effect of the cost matrix in classification on balancing between the true-positive rate and true-negative rate if novel locations are considered as positive. The reason for this study is that the regular locations are usually more accurately predicted so that we may expect that the true-negative rate should be smaller. In particular, we are interested in the effect of the ratio (c) of the cost of classifying a regular location as novel to the cost of classifying a novel location as regular. Its results are manifested in Figures 5(c) and 5(f). It is clear that the larger values of c imply a smaller false-positive rate and true-positive rate. In particular, when the false-positive rate is around 0.1, the true-positive rate using all features and using historical and temporal features (HistTime) can be over 0.6 and 0.4, respectively, in the best case on both datasets. Therefore, in order to serve well for connecting location prediction and location recommendation, this parameter needs to be carefully tuned.

5.2. HMM for Regular Location Prediction

Given the transition probability and emission probability of HMM, we can easily compute the probability of POIs given a previous POI and time (hour of day and day of week) for prediction. In order to study the effect of these two factors, *Markov-S,* which only uses transition probability, and *Regularity,* which only utilizes emission probability, are also used as baselines. The overall comparison on both datasets is shown in Figure 6. It indicates the following observations. (1) Regularity can outperform HOD and HOW as well as MostFreq. Therefore, without any help of smoothing techniques, both HOD and HOW suffer from insufficiency of training data. This result is validated by the lower performance of HOW compared to HOD and is even much lower than MostFreq. (2) A Markov model without any smoothing method is worse than a smoothed Markov model and even worse than MostFreq. Therefore, a Markov model without smoothing also suffers from the insufficiency of training data. (3) A Markov with fallback is worse than HPYP and Markov-S. This is because it doesn't discount the existing observed times of transition but switches to MostFreq when the Markov model encounters a tie problem, where more than two locations have an equal time (including zero) of transition from the previous one. (3) Markov-S is comparable to HPYP, although according to [Teh 2006b], HPYP could be better than Markov-S. The reason is that there are seldom



(a) Gowalla

(b) Jiepang

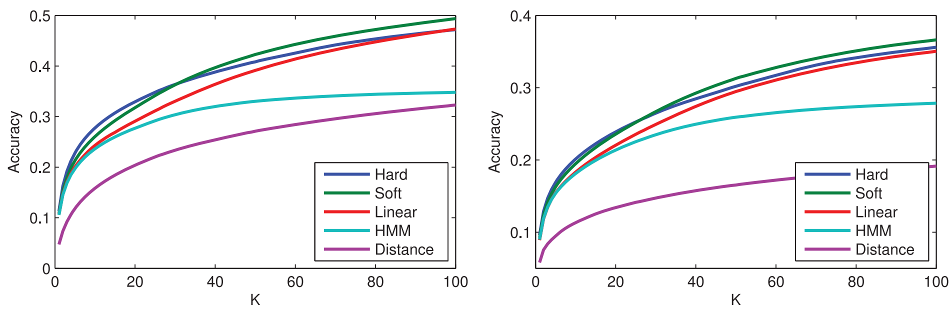
Fig. 7. Novel location prediction performance. "USD" is the combination of UCF, SCF, and density. "All" is the combination of "USD" and "Distance."

frequent transitions across locations, while the major difference between HPYP and Markov-S lies in the different discounts for different transition frequencies. In particular, transitions being observed more than two times only occupy 0.5% on Jiepang and 1% on Gowalla of all transitions. (4) The HMM framework for integrating regularity and a Markov model outperforms each component, but the benefit of the Markov model is limited when compared to Regularity. Nevertheless, the overall prediction accuracy is around 0.1 on both datasets and is far lower than the prediction accuracy (around 0.7) on the continuous trajectories (cell tower records, GPS trajectories). This may be because check-ins are users' voluntarily reported mobility data so that much important data is missing. Although this indicates the difficulty of predicting the next location on location-based social networks, it creates new opportunities for designing novel algorithms for them, which is just the goal of this article.

In addition to the aforementioned experiments, we also study the influence of distance from the immediately preceding locations on the regular location prediction. In particular, in the HMM framework, we are already able to compute the probability $P(l|t,k)$ (t is current time and k is an immediately preceding location, while l is a next possible location), and then the distance $d(l,k)$ is incorporated into HMM as $P(l|t,k)P(d(l,k))$. Here, $P(d(l,k))$ is computed by plugging $d(l,k)$ into the probability distribution obtained by the density estimation on the distance between consecutive check-ins. The experimental results show that introduced distance doesn't significantly improve the performance and thus it is not placed here. However, this result is not surprising since the POIs visited by the same user tend to be clustered geographically according to Ye et al. [2011].

5.3. Recommendation for Novel Location Prediction

The prediction performance of novel location with respect to different configurations of features on both datasets is illustrated in Figure 7, which shows that the combination of User-based CF (UCF), Social-based CF (SCF), and Density (KDE on spatial distribution), that is, USD, could outperform individuals, similar to the results in Ye et al. [2011]. Another result is that SCF generates small but significant benefits, which confirms the discovery in Cho et al. [2011]. However, the performance of Density is not as good as UCF, which is different from Ye et al.'s results. After comparing the density-estimation-based algorithm with theirs, we don't find significant differences in recommendation performance. In this case, this could arise for two reasons. First, the ways of splitting a dataset for training and testing affect the difference between UCF and Density. By conducting experiments using different splitting strategies, one of which splits datasets according to a chronological order and the other of which splits



(a) Gowalla (b) Jiepang

Fig. 8. Performance of CEPR. "Linear" means linear combination of these two models, that is, $P(l) = \alpha P_r(l) + (1-\alpha)P_n(l), \alpha = 0.76$.

randomly using the same percentage (70% as training), we find that in the latter split strategy, Density is better than UCF. This is because when a user is pretty familiar with an area, he or she is less likely to continue location exploration. Therefore, this feature can be very useful for recommendations when a user has just entered an unfamiliar area. Second, the density of POIs affects the performance of the density-based feature since Density on Jiepang performs much worse than UCF. To investigate its effect, we measure for each POI the number of neighbor POIs within 1km. We then find that it is close to 1,000 on Jiepang on average, while it is only around 200 on Gowalla. Another observation made from Figure 7 is that the distance from the previous location becomes an important feature for predicting the next novel location. This coincides with our previous findings in Figures 2(a) and 2(b). Finally, the combination of distance with the other three features ("All") performs best. It indicates that these features complement each other for novel location prediction.

5.4. CEPR for Location Prediction

After studying the performance of both novel and regular location prediction, we now turn to address the superiority of CEPR—the integration algorithms. In addition to the two aforementioned integration approaches, another method that one can easily think of is using a linear combination ("Linear"), which is formalized as $P(l) = \alpha P_r(l) + (1-\alpha)P_n(l)$. The better value of α is selected as 0.76 based on its performance on the validation dataset. Therefore, the performances of three different types of integration are shown in Figure 8. For the sake of better comparison, we also plot the curve of regular location prediction (HMM) and the distance-based ranking algorithm (Distance), which uses the distance from the immediately preceding locations for ranking. This figure shows that both the "Soft" and "Hard" combinations outperform the linear combination. Based on two times of paired two-sample t-test, the superiority of the "Soft" model to "Linear" is significant in all ranking positions, while the superiority of the "Hard" model to "Linear" is only significant in the upper positions ($K \leq 50$). In other words, in the latter case, when compared to the linear combination, the "Hard" one has a particularly better performance in the upper position. This is because in the linear combination, the large value of α renders the dominating effect of $P_r(l)$ to $P_n(l)$. However, due to the existence of both models at any time in "Linear," the advantage of "Hard" over "Linear" gradually becomes smaller with the increasing number of candidates chosen for prediction. Moreover, this is also because the "Hard" approach will be affected by the performance of Exploration Prediction due to its hard assignment. Fortunately, this case can be alleviated to some extent by the soft assignment, which can be considered as the compromise between "Hard" and "Linear." This implies the superiority of "Soft" compared to "Linear" and "Hard," just shown in these two figures. However, their gaps are not extremely large, which encourages us to design a more effective integration method for making the best use of each component. By comparing HMM with CEPR in any case, it is clear that CEPR is much better, improving results by as much as 30%. This indicates that the novel location prediction (recommendation) can serve well for the overall prediction after being integrated with regular location prediction (HMM). Comparing Distance to HMM, Distance is much worse than HMM, although it is important in novel location prediction. This is compatible with the insignificant contribution of Distance to HMM in regular location prediction.

6. CONCLUSIONS

In this article, we studied the Exploration Prediction (EP) problem and boiled it down to a binary classification problem. We then proposed three types of features and trained Logistic Regression (LR) and Classification and Regression Tree (CART) for this problem. They were evaluated on two large-scale check-in datasets from location-based social networks. The experimental results showed that CART outperformed LR due to nonlinearity in EP, achieving around a 20% classification error rate and greatly outperforming the MostFrequent classifier. By studying the importance of different features, we found that current status of novelty seeking (measured as a novel ratio) and users' personality trait of neophilia (measured as users' entropy), as well as the familiarity with the visited location (measured as a visiting ratio), had the largest effect on EP.

Then we proposed the Collaborative Exploration and Periodically Returning (CEPR) model based on EP so that it alternated between a location recommendation model and a location prediction model according to the results of EP. By evaluating them on two aforementioned datasets, we observed that CEPR improved the performance by as much as 30% compared to the traditional location prediction algorithms. Our prediction algorithm in CEPR addressed the problem stemming from the scarcity of individuals' check-in history and the large number of individually visited locations. Particularly, we exploited kernel smoothing techniques for regularity and widely used interpolating techniques in a language model for a Markov model. The experimental results showed the superiority of these smoothed versions compared to nonsmoothed ones. Our recommendation algorithm in CEPR leveraged a learning-to-rank algorithm to integrate users' preferences, social and geographical inferences, and distance from the previous location. This recommendation algorithm, as well as all of its features, was also evaluated on both LBSN datasets, where we observed their effectiveness. Moreover, we also greatly reduced the time complexity of inferring geographical influence based on two-dimensional kernel density estimation.

Until now, we have seen the effect of Exploration Prediction, but the accuracy of Exploration Prediction is still not at a sufficient level. The first possible improvement can be resorting to more complex classification models, such as kernel logistic regressions, which not only have a natural probabilistic interpretation but also can handle nonlinearity of features. The second improvement could be proposing more features, for example, leveraging novelty-seeking preferences of friends or personalized temporal preferences. In addition to making improvements on EP, it is also possible to make better use of its result to integrate location predictions with recommendations, for example, trying different integration strategies in addition to the existing ones.

ACKNOWLEDGMENTS

We thank the anonymous reviewers for their helpful comments and constructive suggestions.

REFERENCES

D. Ashbrook and T. Starner. 2002. Learning significant locations and predicting user movement with GPS. In *Proceedings of ISWC'02*. IEEE, 101–108.

M. Azizyan, I. Constandache, and R. R. Choudhury. 2009. SurroundSense: Mobile phone localization via ambience fingerprinting. In *Proceedings of MobiComp'09*. ACM, 261–272.

J. L. Bentley and H. A. Maurer. 1980. Efficient worst-case data structures for range searching. *Acta Informatica* 13, 2 (1980), 155–168.

C. Burges, T. Shaked, E. Renshaw, A. Lazier, M. Deeds, N. Hamilton, and G. Hullender. 2005. Learning to rank using gradient descent. In *Proceedings of ICML'05*. ACM, 89–96.

J. Chang and E. Sun. 2011. Location3: How users share and respond to location-based data on social. In *Proceedings of ICWSM'11*.

S. F. Chen and J. Goodman. 1996. An empirical study of smoothing techniques for language modeling. In *Proceedings of ACL'96*. ACL, 310–318.

C. Cheng, H. Yang, I. King, and M. R. Lyu. 2012. Fused matrix factorization with geographical and social influence in location-based social networks. In *Proceedings of AAAI'12*.

Z. Cheng, J. Caverlee, K. Lee, and D. Z. Sui. 2011. Exploring millions of footprints in location sharing services. In *Proceedings of ICWSM'11*.

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

K. Dehnad. 1987. Density estimation for statistics and data analysis. *Technometrics* 29, 4 (1987), 495–495.

C. Desrosiers and G. Karypis. 2011. A comprehensive survey of neighborhood-based recommendation methods. In *Recommender Systems Handbook*. Springer, 107–144.

R. P. Ebstein, O. Novick, R. Umansky, B. Priel, Y. Osher, D. Blaine, E. R. Bennett, L. Nemanov, M. Katz, and R. H. Belmaker. 1996. Dopamine D4 receptor (D4DR) exon III polymorphism associated with the human personality trait of novelty seeking. *Nature Genetics* 12, 1 (1996), 78–80.

V. Etter, M. Kafsi, and E. Kazemi. 2012. Been there, done that: What your mobility traces reveal about your behavior. *Mobile Data Challenge by Nokia Workshop, in conjunction with Int. Conf. on Pervasive Computing.*

H. Gao, J. Tang, X. Hu, and H. Liu. 2013. Modeling temporal effects of human mobile behavior on location-based social networks. In *Proceedings of CIKM'13*.

H. Gao, J. Tang, and H. Liu. 2012a. Exploring social-historical ties on location-based social networks. In *Proceedings of ICWSM'12*.

H. Gao, J. Tang, and H. Liu. 2012b. gSCorr: Modeling geo-social correlations for new check-ins on location-based social networks. In *Proceedings of CIKM'12*. ACM, 1582–1586.

H. Gao, J. Tang, and H. Liu. 2012c. Mobile location prediction in spatio-temporal context. In *Proceedings of the Mobile Data Challenge at the 10th International Conference on Pervasive Computing*.

M. Ge, C. Delgado-Battenfeld, and D. Jannach. 2010. Beyond accuracy: Evaluating recommender systems by coverage and serendipity. In *Proceedings of RecSys'10*. ACM, 257–260.

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

T. Hastie, R. Tibshirani, and J. H. Friedman. 2001. *The Elements of Statistical Learning*. Vol. 1. Springer New York.

N. Hurley and M. Zhang. 2011. Novelty and diversity in top-N recommendation–analysis and evaluation. *ACM Transactions on Internet Technology (TOIT)* 10, 4 (2011), 14.

D. Lian and X. Xie. 2011. Learning location naming from user check-in histories. In *Proceedings of GIS'11*. ACM, 112–121.

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

B. Liu and H. Xiong. 2013. Point-of-interest recommendation in location based social networks with topic and location awareness. In *Proceedings of SDM'13*. SIAM, 396–404.

T. Y. Liu. 2009. Learning to rank for information retrieval. *Foundations and Trends in Information Retrieval* 3, 3 (2009), 225–331.

J. McInerney, S. Stein, A. Rogers, and N. R. Jennings. 2013a. Breaking the habit: Measuring and predicting departures from routine in individual human mobility. *Pervasive and Mobile Computing* 9, 6 (2013), 808–822.

J. McInerney, J. Zheng, A. Rogers, and N. R. Jennings. 2013b. Modelling heterogeneous location habits in human populations for location prediction under data sparsity. In *Proceedings of Ubicomp'13*. ACM, 469–478.

A. Monreale, F. Pinelli, R. Trasarti, and F. Giannotti. 2009. WhereNext: A location predictor on trajectory pattern mining. In *Proceedings of KDD'09*. ACM, 637–646.

A. Noulas, S. Scellato, N. Lathia, and C. Mascolo. 2012a. Mining user mobility features for next place prediction in location-based services. In *Proceedings of ICDM'12*. IEEE, 1038–1043.

A. Noulas, S. Scellato, N. Lathia, and C. Mascolo. 2012b. A random walk around the city: New venue recommendation in location-based social networks. In *Proceedings of SocialCom'12*. IEEE, 144–153.

A. Ofstad, E. Nicholas, R. Szcodronski, and R. R. Choudhury. 2008. Aampl: Accelerometer augmented mobile phone localization. In *Proceedings of the fi1st ACM International Workshop on Mobile Entity Localization and Tracking in GPS-Less Environments*. ACM, 13–18.

R. Pan, Y. Zhou, B. Cao, N. N. Liu, R. Lukose, M. Scholz, and Q. Yang. 2008. One-class collaborative filtering. In *Proceedings of ICDM'08*. IEEE, 502–511.

S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme. 2009. BPR: Bayesian personalized ranking from implicit feedback. In *Proceedings of UAI'09*. AUAI Press, 452–461.

A. Sadilek, H. Kautz, and J. P. Bigham. 2012. Finding your friends and following them to where you are. In *Proceedings of WSDM'12*. ACM, 723–732.

G. Shani, D. Heckerman, and R. I. Brafman. 2005. An MDP-based recommender system. *Journal of Machine Learning Research* 6 (2005), 1265–1295.

B. Shaw, J. Shea, S. Sinha, and A. Hogue. 2013. Learning to rank for spatiotemporal search. In *Proceedings of WSDM'13*. ACM, 717–726.

C. Song, T. Koren, P. Wang, and A. L. Barabási. 2010a. Modelling the scaling properties of human mobility. *Nature Physics* 6, 10 (2010), 818–823.

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

L. Song, D. Kotz, R. Jain, and X. He. 2004. Evaluating location predictors with extensive Wi-Fi mobility data. In *Proceedings of INFOCOM'04*, Vol. 2. IEEE, 1414–1424.

M. Szell, R. Sinatra, G. Petri, S. Thurner, and V. Latora. 2012. Understanding mobility in a social petri dish. *Scientific Reports* 2 (2012).

Y. W. Teh. 2006a. *A Bayesian Interpretation of Interpolated Kneser-Ney*. Technical Report TRA2/06. School of Computing, National University of Singapore.

Y. W. Teh. 2006b. A hierarchical Bayesian language model based on Pitman-Yor processes. In *Proceedings of ACL'06*. ACL, 985–992.

W. R. Tobler. 1970. A computer movie simulating urban growth in the Detroit region. *Economic Geography* 46 (1970), 234–240.

S. Vargas and P. Castells. 2011. Rank and relevance in novelty and diversity metrics for recommender systems. In *Proceedings of RecSys'11*. ACM, 109–116.

C. Wang and B. A. Huberman. 2012. How random are online social interactions? *Scientific Reports* 2 (2012).

D. Yang, D. Zhang, Z. Yu, and Z. Wang. 2013. A sentiment-enhanced personalized location recommendation system. In *Proceedings of the 24th ACM Conference on Hypertext and Social Media (HT'13)*. ACM, 119–128.

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

D. Zhang, C. Chen, Z. Zhou, and B. Li. 2012. Identifying logical location via GPS-enabled mobile phone and wearable camera. *International Journal of Pattern Recognition and Artificial Intelligence* (2012).

Y. Zhang, M. Zhang, Y. Liu, S. Ma, and S. Feng. 2013. Localized matrix factorization for recommendation based on matrix block diagonal forms. In *Proceedings of WWW'13*. International World Wide Web Conferences Steering Committee, 1511–1520.

V. W. Zheng, B. Cao, Y. Zheng, X. Xie, and Q. Yang. 2010a. Collaborative filtering meets mobile recommendation: A user-centered approach. In *Proceedings of AAAI'10*. AAAI Press.

V. W. Zheng, Y. Zheng, X. Xie, and Q. Yang. 2010b. Collaborative location and activity recommendations with gps history data. In *Proceedings of WWW'10*. ACM, 1029–1038.

Y. Zheng, L. Zhang, Z. Ma, X. Xie, and W. Y. Ma. 2011. Recommending friends and locations based on individual location history. *ACM Transactions on the Web* 5, 1 (2011), 5.

Received October 2013; revised February 2014; accepted April 2014