Journal of Network and Computer Applications 108 (2018) 64–75

Contents lists available at ScienceDirect

Journal of Network and Computer Applications

journal homepage: www.elsevier.com/locate/jnca

Effective fine-grained location prediction based on user check-in pattern in LBSNs

Jiuxin Cao [*,1], Shuai Xu [1], Xuelin Zhu, Renjun Lv, Bo Liu

School of Computer Science and Engineering, Southeast University, Nanjing, China

ARTICLE INFO

Keywords:
Location-Based Social Networks
Location prediction
Check-in pattern analysis
Scoring model
Classification model

ABSTRACT

Location-Based Social Networks (LBSNs) have built bridges between virtual space and real-world mobility in recent years. The massive check-in data generated in LBSNs makes it possible to predict users' future check-in location, which has proved meaningful for e-commerce developments. Existing studies mainly focus on predicting the next check-in location with a coarse granularity, which only shows limited performance in practical scenarios. In this paper, we propose a comprehensive approach based on user check-in pattern to predict users' future check-in location at any fine-grained time in LBSNs. Firstly, users' check-in pattern involving time periodicity, global popularity and personal preference are analyzed. Secondly, we extract multiple features related to user check-in pattern and explore the predictive power of each individual feature. Thirdly, a set of features are combined into a supervised scoring model and a classification model respectively for predicting user's check-in location at a fine-grained time in the future. Finally, extensive experiments on three real-world Foursquare datasets are carefully designed to verify the effectiveness of the proposed approach. Experimental results show that our approach outperforms both baseline methods and state-of-the-art methods on various evaluation metrics.

1. Introduction

The inflating prevalence of mobile Internet hastens the popularity of **L**ocation-**B**ased **S**ocial **N**etworks (**LBSN**s), among which some typical platforms like Foursquare have become indispensable for Internet users. Different to traditional Online Social Networks (OSNs), LBSNs are featured by location information through which users are able to share check-ins with their friends especially when they visit interesting places (Zheng, 2011). The massive check-in data generated by millions of users in LBSNs can be used to explore the intrinsic pattern of user check-in behavior, and further predict their future check-in location based on their historical check-in records. This application scenario has proved valuable for traffic planning (Gomes et al., 2013), product recommendation (Bao et al., 2012) and disaster warning (Gao et al., 2011), to name a few.

Research on location prediction has been a hot issue over these years. Many earlier studies are mostly carried out based on GPS trajectory data (Ashbrook and Starner, 2003; Gonzalez et al., 2008; Mathew et al., 2012; Sadilek and Krumm, 2012) or user call records (Cho et al., 2011; Wang and Guo, 2014; De Montjoye et al., 2013; Song et al., 2010). Although trajectory data acquired by GPS devices can cover users' complete location information, the deficiencies still exist in that no semantic information (e.g. what activity a user was doing there) is available in such kind of data. Another drawback is that users' personal privacy may be exposed as they have to inform their real-time position. As for location data which is obtained from user call records, it can only provide coarse information that even two geographically near locations can not be distinguished. Compared with GPS trajectory data and user call records, check-in data in LBSNs (such as Foursquare) is more refined with accurate check-in time, coordinates as well as social relationships (Zheng, 2011). Besides, locations in LBSNs are often tagged with semantic information like category which is convenient for the researchers to conduct multi-angle analysis on user check-in pattern. Therefore, developing location prediction algorithm based on huge check-in data in LBSNs ensures the universality when solving this problem.

Many of the recent studies based on check-in data in LBSNs (refer to Section 2) are concerning predicting the next location that a user is going to check-in. These techniques tend to show limited performance in practical scenarios as they cannot deal with situations in the far

* Corresponding author.
E-mail addresses: jx.cao@seu.edu.cn (J. Cao), xushuai7@seu.edu.cn (S. Xu), zhuxuelin@seu.edu.cn (X. Zhu), lvrenjun@seu.edu.cn (R. Lv), bliu@seu.edu.cn (B. Liu).
[1] These authors contribute equally to this work.

https://doi.org/10.1016/j.jnca.2018.02.007
Received 25 October 2017; Received in revised form 2 January 2018; Accepted 6 February 2018
Available online 12 February 2018
1084-8045/

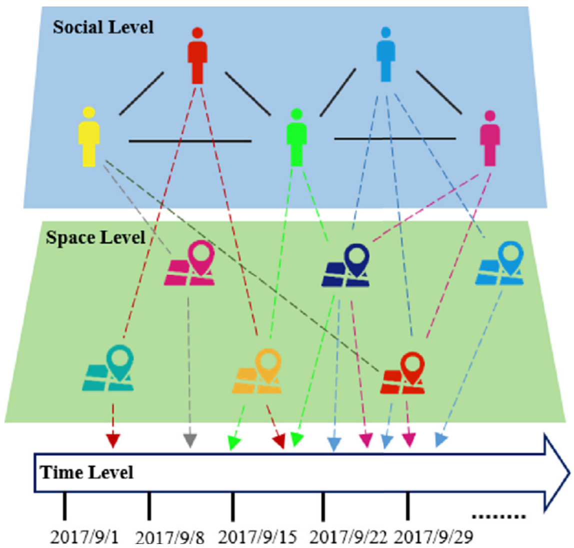

Fig. 1. Framework of the location-based social network.

future. For a few studies that deal with predicting check-in location long in advance, they often pay attention to individual features separately and ignore the integration effect of various factors on user check-in behavior. Besides, existing studies mainly conduct location prediction with a coarse granularity concerning only location category (Kounev, 2012; Likhyani et al., 2015; Ye et al., 2013; Bart et al., 2013) or single day as the time measurement (Zhang et al., 2012). Apparently, the practicability of such location prediction approaches is narrowed in the increasingly fierce competition environment for location-based services.

In view of the shortcomings resided in current studies, we aim to predict users' future check-in location at any fine-grained time accurate to hour based on a comprehensive consideration of time periodicity, global popularity and personal preference. In essence, our goal is to predict users' future check-in probability at a location where they have visited before. To fulfill this task, we firstly analyze user check-in pattern from time periodicity, global popularity as well as personal preference. Then, we combine all these factors into a supervised scoring model and a classification model respectively to solve this problem from two different angles. As is validated in the experiments based on three real-world Foursquare datasets, the proposed approach can predict users' future check-in location with *APR* up to 0.866 and *F*1 score up to 0.777. Besides, it outperforms both baseline methods and sate-of-the-art methods with higher accuracy.

The main contributions of this paper are summarized as follows:

- We analyze user check-in pattern and explore the predictive power of twelve individual features involving time periodicity, global popularity and personal preference on three real-world Foursquare datasets. The observation of both *APR* metric and *Accuracy@N* metric indicates that both time periodicity and personal preference play significant roles among all the individual features.
- We combine all the features into a supervised scoring model to evaluate the possibility of a given user's visit to a candidate location at a given time, thus we can aggregate the features to improve prediction performance. By means of stochastic gradient descend algorithm, we manage to infer the parameters in the scoring model.
- We reduce the location prediction problem to a binary classification task and train two different models to classify whether a given user would check-in at a candidate location at a given time.
- We verify the performance of the proposed approach through extensive experiments on three real-world datasets, and demonstrate its superiority over state-of-the-art methods.

The remainder of this paper is organized as follows. Section 2 reviews the related work. In Section 3, the formal definition of location prediction problem in this paper is briefly described. Section 4 introduces the datasets used for experiment. In Section 5, the predictive power of individual features is explored. Section 6 and Section 7 illustrates the main ideas of the scoring model and classification model respectively, followed by experimental evaluation in Section 8. Finally, Section 9 concludes the paper.

2. Related work

2.1. Predicting next check-in location

Existing studies are mainly about predicting the next location where a user is going to check in. Lian et al. (2014) thoroughly analyze the properties of check-in traces together with users' demographics. The correlation between location predictability and these two aspects shows that users' next check-in place can be well estimated. Gao et al. (2012) propose a comprehensive language model to describe the generation mechanism of the next check-in location. Noulas et al. (2012) study the ability of single user's check-in pattern for predicting the next check-in location and further integrate a set of features together to improve the prediction performance. In order to improve the performance of next location prediction in LBSNs, Likhyani et al. (2015) explore coarse-grained venue categories by associating sparse location data with map information, and then leverage this auxiliary information for more accurate location prediction of the next visit. A similar work is conducted by considering social friendship in (Wang et al., 2017). Based on the assumption that some inherent patterns from past check-ins are capable of indicating future check-in behavior, Kounev (2012) develops two predictors to predict both the next visited venue's category and the expected time of that check-in. Based on the observation that users tend to check-in several times in a single place, a predicting model based on Markov chain is proposed in (Li et al., 2012).

As is mentioned in Section 1, compared with any-time location prediction applications, merely predicting the next check-in location shows limited performance in practical scenarios.

2.2. Predicting far future check-in location

There exists a few studies concerning location prediction long in advance. Pang and Zhang (2015) demonstrate that users' mobility is constrained by the social impact which is reflected through a small subset of his communities. Considering the community information of a user, the method they propose is able to answer the question whether he will check-in at a certain place for a given time. Different to our work in this paper, they ignore the effect of time periodicity which could also play an important role in determining a user's future check-in. Assam and Seidl (2014) quantify the effect of social influence on a user's check-in by wavelets so that the scenarios with more than one influencer are properly scrutinized. Again, the time periodicity is not explored. Cho et al. (2014) observe that check-ins are characterized by temporal dynamics and they sometimes exhibit bursty patterns. Based on this observation, they use the Hawkes process to describe check-in dynamics so that the likelihood of a particular event in the future can be measured by past events. Similar to our work, studies in (Zhang et al., 2012; Bart et al., 2013) extract effective temporal features to learn user check-in pattern from historical check-in records, but the shortcoming lies in that these approaches only provide a coarse granularity concerning venue category or single day as the time measurement.

Compared with the aforementioned studies, our research is featured by a comprehensive analysis on time periodicity, global popularity and personal preference in user check-in behavior, and our proposed approach can provide location prediction accurate to hour, which is promising in the fierce competition for location-based services.