Neurocomputing 176 (2016) 98–106

Contents lists available at ScienceDirect

Neurocomputing

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

Exploiting machine learning techniques for location recognition and prediction with smartphone logs ☆

Sung-Bae Cho

Department of Computer Science, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul 120-749, South Korea

ARTICLE INFO

Article history:
Received 10 March 2014
Received in revised form
29 January 2015
Accepted 6 February 2015
Available online 8 May 2015

Keywords:
Location recognition
Location prediction
Location-based services
Hidden Markov model
k-Nearest neighbor
Decision trees

ABSTRACT

Due to the advancement of mobile computing technology and the various sensors built in the smartphones, context-aware services are proliferating to everyday life. Location-based service (LBS), which provides the appropriate service to smartphone users according to their contexts, is becoming more popular, and the location is one of the most important contexts in LBS. Extracting and recognizing meaningful location and predicting next location are crucial for successful LBS. Many researchers have attempted to recognize and predict locations by various methods, but only few consider the development of real working system considering key tasks of LBS on the mobile platform. In this paper, we propose a location recognition and prediction system in smartphone environment, which consists of recognizing location and predicting destination for users. It recognizes user location by combining k-nearest neighbor and decision trees, and predicts user destination using hidden Markov models. To show the usefulness of the proposed system, we have conducted thorough experiments on real everyday life datasets collected from 10 persons for six months, and confirmed that the proposed system yielded above 90% of average location prediction accuracy.

1. Introduction

As the mobile devices are popular in everyday life, smartphone has become a powerful platform for mobile context-aware services, which have attracted great attention and propelled active investigation about inferring user's mobile contexts for useful services [1–4]. It provides with a lot of interesting information about users, especially as various sensors have been equipped in recent mobile devices. One of the most important contexts might be the location. The proper services and information could be delivered according to user's current location or future location. In spite of a lot of research with respect to location prediction, however, there are few real working systems which recognize and predict the location on the mobile device.

In this paper, we address the key issues of location-based services by exploiting machine learning techniques to develop a hybrid AI system for location prediction with smartphone logs. The main components of the system are as follows.

1.1. Extracting meaningful locations

To effectively extract the meaningful locations, we exploit the G-means clustering method, which automatically determines the number of clusters by performing statistical test iteratively. The stay points are identified automatically for G-means algorithm to reduce the burden in the mobile computing.

1.2. Recognizing and predicting locations

To effectively recognize the locations in outdoor as well as in indoor, we utilize several classification methods such as k-nearest neighbor (kNN) and multiple decision trees (DT). The location prediction problem is formulated as a path classification and model selection problem. The proposed method constructs a hidden Markov model (HMM) for each path by utilizing intermediate locations, which come from the location recognition phase, and predicts the next location by selecting the model that produces the largest matching score. The model uses the smartphone logs such as transportation mode, day of week and time to predict the context-sensitive locations reliably.

1.3. Managing models and log data

To make the system user-friendly, we develop an intuitive user interface for location recognition and prediction. The system has

☆This is the extended paper presented at the HAIS: Y.-J. Kim and S.-B. Cho, A HMM-based location prediction framework with location recognizer combining k-nearest neighbor and multiple decision trees, in: Proceedings of the International Conference on Hybrid Artificial Intelligence Systems, pp. 618–628, 2013.

E-mail address: sbcho@cs.yonsei.ac.kr

http://dx.doi.org/10.1016/j.neucom.2015.02.079
0925-2312/

the functions of location management and path management to manage user's locations and paths appropriately.

The problem of location prediction can be formulated as follows. Given an observed user trajectory $T=\ell_0\rightarrow\ell_1\rightarrow\ell_2\rightarrow\cdots\rightarrow\ell_t$ and the location set $L=\{L_1, L_2, L_3, \ldots, L_n\}$, where ℓ_i means the ith location that user has visited, we want to predict the next location ℓ_{t+1}. In Fig. 1, given the information of trajectory, $L_2\rightarrow L_4\rightarrow L_3\rightarrow\ldots\rightarrow L_5$, we have to predict the next significant location of L_4.

In this paper, the location prediction problem is transformed into the path classification problem. Given an observed user trajectory, $T=\ell_0\rightarrow\ell_1\rightarrow\ell_2\rightarrow\cdots\rightarrow\ell_t$, the set of locations $L=\{L_1, L_2, L_3, \ldots, L_n\}$, the set of user's movement paths $P=\{p_1, p_2, p_3, \ldots, p_m\}$ where $p_i=(\ell_s^i, \ell_d^i)$ (s and d denote the start and destination indicators, respectively), and the set of intermediate locations $I=\{I_1, I_2, I_3, \ldots, I_n\}$ between the start and destination locations of each path p_i, we have to classify an observed user trajectory T into one of the user's movement paths. By solving this problem, we can get a classified path and predict user's next location by returning destination location of the path.

Many researchers have attempted to work out the problem of location recognition and prediction by using several approaches. Table 1 shows some related works on location prediction.

In addition, Ashbrook et al. extracted user's significant locations from GPS data and presented a location predictor based on the Markov models [11]. Krumm et al. devised a method called predestination that predicts driver's destination as trip progresses [12]. Alvarez-Garcia et al. presented a new approach to predict destinations given the only data of a partial trip by using HMMs and local street-map [13]. Petzold et al. presented a dynamic Bayesian network to predict the next indoor location and compared with the state predictor and multi-layer perceptron predictor [14].

Yavas et al. presented a data mining algorithm for the prediction of user movements in a mobile computing system [15]. Monreale et al. proposed the trajectory pattern tree aimed at predicting the next location of a moving object with a certain level of accuracy [16]. Morzy mined the database of moving object locations to discover frequent trajectories and movement rules, and matched the trajectory of a moving object with the database of movement rules to build a probabilistic model [17]. Petzold et al. compared various methods for the next location prediction [18]. They conducted the comparative experiments using dynamic Bayesian network, multi-layer perceptron, Elman net, Markov predictor and state predictor. Scellato et al. presented a novel framework for predicting user's next locations based on nonlinear time series analysis [19].

Most of the works related with location prediction have the limitation that they mainly focus on the performance of methods and only few consider the development of real working system on mobile computing environment. In this paper, we present a hybrid system of location recognition and prediction with smartphone and conduct a realistic experiment with large data collected by 10 users for six months. The system can be used personally, so that the number of users is less important than the duration of collection. We have collected the everyday life data for six months per each user, which is not trivial and expensive to obtain.

The rest of this paper is organized as follows. Section 2 presents the details of the proposed system. Section 3 illustrates the personalized location prediction system implemented and the results of experiments. The summary and future work are described in Section 4.

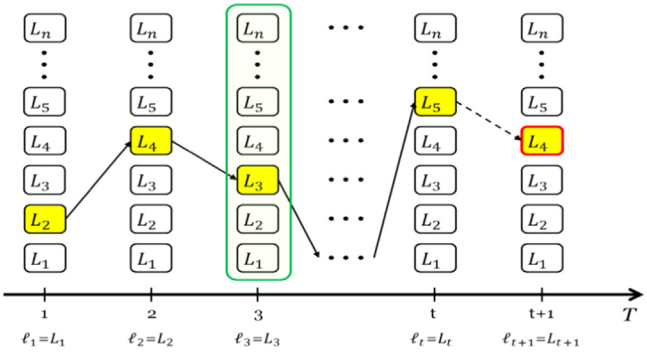

Fig. 1. An example of location prediction problem.

2. The proposed system

The proposed system for predicting user's next location is illustrated as shown in Fig. 2. k-Nearest neighbor (kNN) and decision tree (DT) are used for recognizing the current location from the information of T_n and S_n; the time and transportation mode when user visits the nth location, respectively. The information can be used to model user's paths discriminatively. Hidden Markov models (HMMs) are used for predicting the destination by path classification.

2.1. Location extraction

To model user's paths effectively, we extract the intermediate locations between start and destination locations by clustering GPS data. Most of the previous works based on clustering method used k-means clustering algorithm [11,20,21]. However, this is not

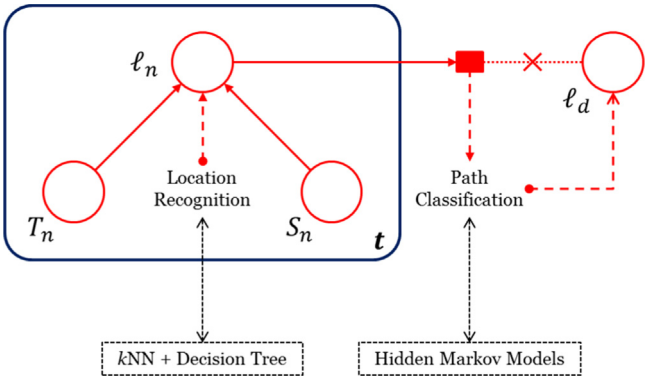

Fig. 2. Overview of the proposed system.

Table 1
Related works of location prediction.

Year	Author	Input	Method	Dataset
2012	Mathew et al. [5]	Location sequence, time, day of week	HMM	GeoLife dataset
2012	Gambs et al. [6]	POI sequence	Mobility Markov Chains	Phonetic dataset, GeoLife dataset, Synthetic dataset
2012	Heo et al. [7]	GPS, activity, day of week, time, velocity	Hierachical DBN	Data for 2 months
2008	Burbey et al. [8]	Location, time	PPM	UCSD dataset
2006	Simmons et al. [9]	GPS, Map DB	HMM	Data for 1 month
2004	Liao et al. [10]	GPS	Hierarchical DBN	Data for 2 months