A comprehensive survey of recommendation system based on taxi GPS trajectory

Yuanhang Hu, Yujiu Yang
Shenzhen Key Laboratory of Broadband Network & Multimedia,
Graduate School at Shenzhen, Tsinghua University
Shenzhen, China
huyh12@mails.tsinghua.edu.cn
yang.yujiu@sz.tsinghua.edu.cn

Biqing Huang
Department of Automation
Tsinghua University
Beijing, China
hbq@mail.tsinghua.edu.cn

***Abstract*—In recent years, the service system based on taxi GPS trajectory gradually becomes a hot research topic. In this paper we first give some descriptions and definitions of the taxi GPS trajectory problems. Different from the traditional recommendation system, the service system based on taxi GPS trajectory will lead to some special challenges and we will give corresponding solutions to solve these problems. Therefore, we propose a recommendation system framework for this issue via the emphasis on temporal and spatial information mining. Then, we discuss the different classification method by different points of views including the statistics of spatial information, the modeling of time information, mining methods and knowledge discovery models. Finally, we point out the promising directions in this field.**

Keywords- GPS trajectory; recommendation system; data mining; survey

I. INTRODUCTION

Taxi service plays an important role in public transportation by offering passengers quick personalized destination service [1]. However, this service is often inefficient when taxi drivers suffer from heavily congested traffic and longer cruising miles. Sometimes taxi drivers need recommendation systems to guide their services instead of using their own experience.

GPS has become a powerful ubiquitous sensor in our daily life. In modern cities, taxis have been equipped with GPS devices [2]. In the beginning, GPS devices were mainly used for localization and navigation. As the number of taxis which are equipped with GPS sensors, large amount of taxi trajectory can be collected in real-time. These traces can be very helpful for improving taxi services.

By mining the GPS trajectory and leveraging the traffic information, recommendation systems based on GPS trajectory play a vital role for taxi service. This survey present a panorama of the service system based on GPS trajectory, facilitating research into this rising topic. The contributions of this article are detailed as follows:

- We propose a general framework with emphasis on temporal and spatial information mining
- We summarize the major method lies used for the systems and discuss their advantages and disadvantages.
- We point out promising research directions in the service system.

The remainder of the paper is organized as follows. In Section II, we introduce basic concepts and give definitions of different problems. In Section III, we propose a general framework for the recommendation system and discuss detailed methodologies from multiply perspectives. In Section IV, we point out the promising research directions for the mobile recommender system. In Section V, we give our concluding remarks.

II. PROBLEM DESCRIPTIONS

A. Basic concepts

To facilitate the presentation of the taxi GPS data mining applications, we define some generalized terms for basic concepts.

GPS data is a tuple: taxi license, longitude, latitude, timestamp, ID, GPS speed, GPS bearing, GPS state, occupied state, detailed in Table I.

TABLE I. SAMPLE OF THE TAXI GPS DATA

data	**Taxis**		
	Taxi 1	***Taxi 2***	***Taxi 3***
Taxi license	粤 B494G3	粤 BN4M66	粤 B207G5
Longitude	113.800720	114.063049	113.836914
Latitude	22.678482	22.533800	22.727600
Timestamp	2013-11-17 00:02:09	2013-11-17 00:02:05	2013-11-17 00:02:05
ID	1090010	1247262	1081595
GPS speed	1	17	25
GPS bearing	183	343	0
GPS state	0	0	0
Occupied state	0	0	1

Based on the origin GPS data, the researchers usually adopt the following concepts:

A trajectory *T* is a sequence of GPS data by a taxi. Its attributes are listed in Table II.

TABLE II. ATRRIBUTIONS OF TRAJECORY AND ROAD SEGMENT

Trajectory	**Attributes**	**Road segment**	**Attributes**
T.id	ID	r.start	Start
T.time	Timestamp	r.end	End
T.state	Occupied state	r.dir	One-way or bidirection

2165-3828/15 $31.00 © 2015 IEEE
DOI 10.1109/ICSS.2015.31

T.longitude	Longitude	r.speed	speed
T.latitude	Latitude	r.way	Lane numbers
T.speed	GPS speed		
T.gpsdir	GPS bearing		
T.validity	GPS state		

According to the T.state, the state of a working taxi can be divided into three kinds: Occupied (O), Cruising(C), and Waiting (W), listed in Table III.

TABLE III. THE STATES OF A TAXI

states	status	Description
Working	Occupied(O)	Known from T.state
Non-working	Cruising(C)	Distinguishing traffic jams when cruising and really waiting places
	Waiting(W)	

According to the daily operation we can get the transformation map of states Fig.1.The taxi usually drop off the passengers at certain location, cruises for a while and picks up some potential passengers and this goes on.

A road segment *r* is a directed edge that forms parts of the road network, detailed in Table II.

A route *R* is a sequence of connected segments, i.e., R: $r_1 \rightarrow r_2 \rightarrow \ldots \rightarrow r_n$, where r_1.end=r_2.start.

A taxi trip t is a sub-trajectory when the taxi is occupied, i.e., {t∈R, t.state=Occupied}. Sometimes t.start is called origin and t.end is called destination.

A hotspot *H* or parking place is one of popular places where taxis often wait for passengers.

A landmark *L*[8] is one of road segments that are frequently traversed by taxi drivers.

B. *Problem definitions*

In specific applications, for different target the recommendation system can be divided into the passenger-based and taxi-based recommendation (the latter is the mainstream); for different guidance level, the recommendation system can be divided into the operational and strategic recommendation. (The former is the mainstream).The detail is listed in Table IV.

TABLE IV. ATRRIBUTIONS OF TRAJECORY AND ROAD SEGMENT

Target	Guidance level	Typical problems
passenger	Operational	Waiting place recommendation
Taxi drivers	Operational	Single hotspot recommendation
		Moving sequence recommendation
		Hotspot recommendation with rountes
	Strategic	Macro passenger-finding stragy

1) Problem 1: Waiting place recommendation.

Given GPS datasets, recommending the place where the passenger should go to take taxis. Normally, it is converted into a optimal prediction of hotspots. The criterion is the possibility of taking a taxi in an given time or the expected waiting time.

2) Problem 2:Single hotspot recommendation

Given GPS datasets, recommending a potential place for passengers. It can also be converted into an optimal prediction of hotspots.

Math description: Given the current position and time, finding out a set of potential hotspots $H = \{h_1, h_2, \ldots, h_N\}$. The goal is to calculate $\arg\max_j f(h_j, p_t)$.

3) Problem 3: Moving sequence recommendation

It is the expansion of Problem 3 with the requirement that the taxi should traverse among the hotspots.

Math description: Given the current position and time, finding out a set of potential hotspots $H = \{h_1, h_2, \ldots, h_N\}$ and forming different moving sequences $H_i : h_{i1} \rightarrow h_{i2} \rightarrow \ldots \rightarrow h_{in}$. The goal is to calculate $\arg\max_j f(H_j, p_t)$.

Research on this problem focuses on the discussion of optimal conditions. [10] regards optimal conditions as PTD(Potential Travel Distance) but [11] thinks MNP(Maximum Net Profit) is more reasonable.

4) Problem 4: Hotspot recommendation with rountes

Problem 2 and Problem 3 focus on the discussion of hotspots or traversing among hotspots, while Problem 4 considers that the routes connecting these hotspots also have profit.

[4] took the profit of each road segments into calculation and implemented a system called HUNTS to give the profit maximizing routines.

5) Problem 5:Macro passenger-finding strategise

The macro strategies involves hunting for passengers and waiting for passengers. One typical discussion is [2]. They built a tuple of Times-Location-Strategy and used L1-Norm-SVM as a feature selector to discover the efficient/inefficient passenger-finding strategies.

In order to solve the problems above, we take a further step to decompose the big recommend problems into three important sub-problems.

1) Clustering of hotspot

The hotspot is the place where taxis frequently wait for passengers. The clustering results directly impact the final recommended results [14][17].

2) Prediciton of traffic status

The traffic status refers to a number of measurements to indicate the road traffic. Measurements include the number of empty within the grid, the average road speed, the number of passengers. Accurate prediction of traffic status is beneficial for the profit calculation and the path selection [15][16].

3) Planning of routes

Planning of routes is based on the prediction of traffic status. In fact people often use traffic measures as a weight of routes to solve optimal problems when traversing the whole network [9][12].

C. *Usage and characteristics of the trajectory in the recommended system*

We treat different types of trajectory with different focuses.

1) The orgin and destination o f a trip

The extracted OD (Origin-Destination) points reflect the activity of the location citizens. More precisely, the origin points can be used to predict the demand and the destination can be used for exploring the functional areas [12].

2) Trajectory produced by the working taxi

When the taxi is occupied, the driver usually takes the passengers to the destination as soon as possible. Thus the analysis of its trajectory focuses on the extraction of operational information including the speed and flow status.

3) Trajectory produced by the non-working taxi

When it is unoccupied, the taxi is not necessarily moving in a high speed for potential passengers. The analysis of its trajectory focuses on the pattern analysis of passenger-finding behavior.

The application recommender system based on GPS trajectory has the following features:

1) Limits of predictibility

GPS data is used to record the daily information of taxis' daily operation which involves so many complicated factors. Although research dig some patterns of traffic but essentially the traffic remains unpredictable.

2) Data sparsity

The volume of data collected is huge but according to 2-8 rules there still exists the problem of imbalanced data for the city wide analysis. So we will face the problem of data sparsity when modeling the whole system.

3) Sampling and trajectory judgement problems

There exist some problems in the mobile recommender system for daily use. For example, the sampling rate of the GPS device is not that high because of the power consumption. Common sampling interval is between one minute to three minutes, which brings challenges for the detection of actual trajectory and other information.

D. Difference with the traditional recommender system

After analyzing the characteristics of mobile recommender system based on GPS trajectory, we can see its great differences from the traditional recommender system.

1) Different methods for recommendation

In traditional recommender system [17][18], the core study object is rating matrix. Based on the rating matrix people design kinds of algorithms to add useful information [20][21]. With the rapid growth in the number of users and items, recommender system must have the ability to handle large-scale data, and academics began to design algorithms of low computational complexity requiring high accuracy at the same time. With the in-depth research goes on, researchers have begun to focus on the use of social networks and other useful information.

In mobile recommendation system based on taxi GPS trajectory, the standard data format is a tuple (ID, longitude, latitude, timestamp, taxi status). It is hard to convert into forms of rating matrix, so the traditional method is difficult to use directly. Instead, some machine learning methods appear more often.

2) Different use of context imformation

Context is a multifaceted concept describing important information related with time and space.

In tradition recommender system, context is a small part of the system and can be use to filter data or train the matrix. But for mobile recommendation system, the application is always location based. The trajectory records contain large amount information of time and geographic space and their dependent relationship, which is a key part of study.

3) Different evaluation methods

In fact, there are a variety of evaluation indicators for traditional recommender system such as precision, diversity, novelty, coverage and so on. These indicators have their own advantages and disadvantages for certain situations.

The precision indicator is a popular index for recommendation and can be divided into two situations. The first one is to calculate the difference between the training matrix and the test matrix, such as MAE (mean absolute error) and RMSE (root mean square error).The other one is to evaluate the result base on the top-N recommendation such as NDCG (normalized discounted cumulative gain).

However, the mobile recommender recommendations lack standard evaluation indictors because it solves ia wide range of problems. Sometimes it needs to borrow indictors from other fields.

III. FRAMEWORK AND METHODOLOGY FOR RECOMMENDATION SYSTEM BASED ON GPS TRAJECTORY

A. System Framework

Dealing with GPS data, the system has two distinct characteristics: the first one is the properties of spatial information and the other is the time attribute information. For the standardization of description, we integrate the current work of the various researchers and propose a common framework for general data processing, as shown in Fig.1.

In the framework, the original input data is GPS logs collected by on-board sensors. According to the extraction of the different spatial and time information, the system will use different data mining methods. After that knowledge discovery model is constructed, and then give the results of targeted strategies with the evaluation process.

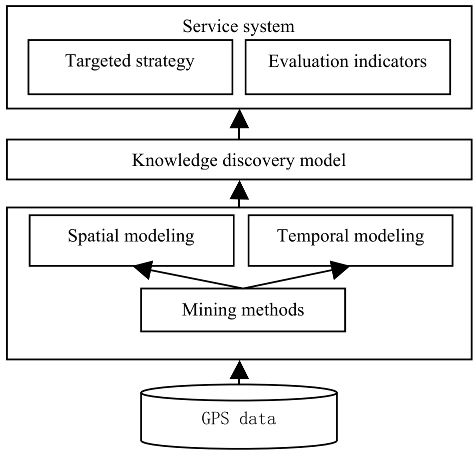

Figure 1. system framework for the service system

B. from the prespective of spatial modeling

Spatial information deals with the location information or the geographic attribute of GPS data. There are two methods handling the spatial information: the grid-based method and the map-matching-based method.

1) the grid-based method

The grid-based method suggests that the geospatial space is divided into grids by latitude and longitude. The advantage is that the model is relatively simple. We only need to considerate the spatial distribution of each single grid and statistical processing would run very fast.

Based on the grid method, [7] calculated pick-up quantity within each grid, and then solved the prediction problem; [6] designed a spatial-temporal profitability map for recommendation and discussed the effect of the grid size.

2) the map-matching-based method

Map matching is the process to match an original GPS tracking data to a digital map or a digital road network. Algorithms for map matching itself have been a hot topic due to different challenges like the low sampling rate problem mentioned above.

The map-matching-based method require the basic map matching function, which means the first step is to do the map matching work [27][28]. Its advantage is clear that we directly simulated trajectory segment and the full amount of information. In fact more systems prefer the map-matching-based method on this basis and it is easy to add other geographic factors, such as landmarks, hotspots, moving sequences, etc.

C. From the perspective of temporal modeling

As we know, GPS data has obvious timing characteristics. The time parameters is important for modeling and are consist of the time point T and the time interval △T.

1) Discussion about T and△T

When time point T serves as a parameter, people intuitively set the value of T as an exact time of the day, such as 7 a.m. When the time interval △T serves as a parameter, people tend to △T is a constant and set the value of △T empirically or through several experiments with discussions.

[6] set a delay experienced window when using a spatial-temporal profitability map and discuss the effect of △T.

2) Learning the time interval

When estimating travel time between the two landmarks, [5] believes that the travel time is not a single value based on time or subject to typical Gaussian distribution, but like a series of set of clusters. So they used Variance-Entropy-Based clustering algorithm to study the temporal distribution. The first step is V-clustering, clustering the travel time between two points and the second step is the E-clustering, using the maximum entropy to partition the time periods.

D. From the perspective of data mining methods and evaluation of the advantages and disadvantages

According to the extraction of the different spatial and time information, we will use different data mining methods for knowledge discovery. These methods include clustering, classification and regression prediction.

1) Clustering for GPS data

Clustering is to divide the data object into several categories, requiring that the object similarity within class is as high as possible and the object similarity between classes is as low as possible. In mobile recommendation, clustering analysis is often used to extract similar structures of trajectory especially in the handling of the clustering of hotspots problem. In addition, we can cluster similar road segments when facing with the data sparsity of different road. The following clustering algorithms are often used: K-means clustering algorithm, hierarchical clustering, density-based clustering and grid-based clustering [30].

[10] extracted a group of points of high performance taxi drivers, applied the K-medoids clustering method and used the center point of each class as a hotspot. [12] used a density-based clustering method called OPTICS (Ordering Points to identify the clustering structure) to find parking spots. [30] selected a grid-based clustering method, including the merging step and the splitting steps, to ensure the effectiveness of the stops.

The advantage of K-mean algorithm is its efficiency and intuition, but the disadvantage is that the selecting of the K value and the clustering result is sensitive to noise; the density-based clustering algorithm is the most common used, and can effectively avoid noise interference, but the difficulty lies in controlling the density of the radius and too dense the area may not be ideal for solving practical problems; hierarchical clustering based on grid can ensure different constraints on the final clustering results and the experiment works best.

2) Classification for GPS data

[12] distinguished the real parking places from the points caused by traffic congestion using bagging classifier and use an artificial data set as a reference marker.

[16] intended to discover both efficient and inefficient passenger-finding strategies for the taxi driver. These strategies are divided into four categories: whether to wait or not and whether to hunting or not. They design a Time-Location-Strategy feature triplet and use L1-Norm SVM as

a feature selection tool and a classifier to make the prediction.

In fact, when processing with the trajectory data, classification algorithms are far less used than clustering algorithms because classification algorithms need to know labels first. But in the complex traffic problems there are few classification applications. Sometimes manual annotation is required.

3) Regression prediction for data

As mentioned above, prediction of traffic status is one of three important sub-problems. Regression prediction models discussed here is often used to solve this problem.In fact, how to extract effective traffic conditions from the GPS data with effective model is a key topic in many related research areas.

a) Naive Bayesian model

Bayesian classification model is a typical method of classification based on statistical models. Bayes theorem is the most important formula in Bayesian theory which describes the relationship between prior probability and posterior probability, as in (1). Priori information and samples of data can be used to determine the posterior probability of certain event.

$$P(Y=y_i|x_1,x_2,...x_n)=\frac{P(Y=y_i)P(x_1,x_2,...x_n|Y=y_i)}{P(|x_1,x_2,...x_n)} \quad (1)$$

Naive Bayes classification model is one of the most simple and effective classification among Bayesian models with basic assumption that the variables are independent and using the principle of MAP (maximum a posteriori) in predicting, as in (2).

$$y_{MAP}=\arg\max_{y_i\in Y} P(Y=y_i)\prod_{i=i}^{n}P(x_i|Y=y_i) \quad (2)$$

Marco Veloso et al [11] use Naive Bayes model to predict the number of empty taxi number Y, taking weeks, hours, and weather as independent variables, as in (3) and (4).

$$P(Y=y_i|week,hour,weather)=\frac{P(Y=y_i)P(week,hour,weather)}{P(week,hour,weather)} \quad (3)$$

$$y_{MAP}=\arg\max_{y_i\in Y}\{P(Y=y_i)P(week|Y=y_i) \quad (4)$$
$$P(hour|Y=y_i)P(weather|Y=y_i)\}$$

The advantage of Bayesian prediction model is its simplicity to use, but the model suffers from a lower accuracy because of the independent assumption and limited selection of independent variables.

b) time series analysis

Time series analysis is widely used in many areas for prediction such as the stock market and can also be introduced to help solve the problem of traffic forecasting. Based on time-series data obtained by systematic observation, time series analysis is to establish the mathematical model by curve fitting and parameter estimation.

Through the GPS data processing, it is easy to get a set of time series Y $\{Y_t,\ t=t,\ t+\triangle t,\ t+2\triangle t\ ...\}$.$\triangle t$ is selected according to the actual need, then time series analysis begins.

ARMA model, short for autoregressive moving average model, is the most commonly used model fitting stationary series ARMA (p, q).

This model includes auto regression model AR (p), as in (5), and moving average model MA (q), as in (6).

$$y_t=\gamma_1 y_{t-1}+\gamma_2 y_{t-2}+...+\gamma_p y_{t-p}+\varepsilon_t \quad (5)$$

$$y_t=\varepsilon_t-\theta_1\varepsilon_{t-1}-\theta_2\varepsilon_{t-2}-...-\theta_q\varepsilon_{t-q} \quad (6)$$

ARIMA model, short for autoregressive integrated moving average model, is seen as an extension of ARMA when the time series itself is not stable, and may become stationary time series after d-order differential, as in (7).

$$\Phi(B)\nabla^d x_t=\mu+\Theta(B)\varepsilon_t \quad (7)$$

[7] applied time series analysis using ARIMA model to derive the internal laws from historical data and predicted the future value. Moreover, they proposed an improved method to obtain better results.

The advantage of ARIMA model is that it combines different time models and is quite applicable for a wide range of problems, but the difficulty lies in smoothing method selection and parameter adjustment.

c) Markov chain theory

A Markov process is a stochastic process with the Markov property which means the next state depends only on the current state and not on the sequence of events that preceded it. the Markov process with the discrete time and states discrete is called a Markov chain. To simplify the problem, people assuming traffic conditions have Markov properties, and build a Markov chain for the predication, as in (8). (x represents a number of history values, y represents the predicted value.)

$$P(Y_n|Y_{n-1}=x_{n-1},Y_{n-2}=x_{n-2},...Y_1=x_1)=P(Y_n|Y_{n-1}=x_{n-1}) \quad (8)$$

The procedures for first order Markov:

- Determine each state space.
- The initial distribution of S (n)
- Calculate one-step transition matrix P(1)
- Calculate S (N) * P $(1)^n$ and use the n-step transfer value for prediction

[5] used the m-order Markov model for predicting traffic conditions, as in (9).

$$P(Y_n|Y_{n-1}=x_{n-1},Y_{n-2}=x_{n-2},...Y_1=x_1)=P(Y_n|Y_{n-1}=x_{n-1},y_{n-2}=x_{n-2},...,Y_{n-m}=x_{n-m}) \quad (9)$$

Markov models can be used for short-term prediction and produce better experimental results than the time-series analysis.

IV. FUTURE WORK AND DIRECTION

After the problem discussions and techniques analysis based on the above framework, we believe that recommendation system based on taxi GPS trajectory mining has the following development directions:

A. the transformation of traditional recommendation

Previous work tends to work on GPS data mining, and rarely see the application of GPS data from the perspective of recommendation system. In fact, the concept of the recommendation system based on taxi GPS trajectory is quite novel. It is natural to compare the new recommendation system with the traditional one. After figuring out the similarities and differences, it is possible to make the transformation of traditional recommendation for mobile recommendation.

B. more refined modeling and personalization

We should build models at a more refined granularity. For example, most of the existing recommender systems for taxis is based on global location, so it is natural to consider the competition if taxi drivers go to the same place.

Personalization is an important indicator of recommendation system. The recommender system have little for the taxi's personality preferences, so giving a more personalized recommendation is one of the important research direction.

For the problem of clustering hotspots, the chosen of clustering algorithms tend to be application based with different constrains such as minimum points within a cluster or required sharp of clustering. These popular clustering technique discussed are only considering the static information of points. It is worth considering using mobility patterns or dynamic information as a mobility-based clustering method for better results.

C. Visualization and online computing

Data visualization has always been an important aspect of the practical application of data mining. Kinds of flexible ways of visualizing recommendation systems based on GPS data containing the temporal and spatial information is also one of the key considerations.

Faced with massive data, the application-level system should consider the offline and online calculation ability and implement algorithms that are required to return the result within an allowable time.

V. CONCLUSION

With the wide use of GPS data, the recommendation application based on taxi GPS trajectory gradually becomes a hot research topic. In this paper, we give a comprehensive survey of recommendation system based on taxi GPS trajectory and propose a general framework for better analysis. We put emphasis on temporal and spatial information mining. Then from the perspective of data mining methods we discuss the three basic techniques including clustering, classification and regression prediction and point out their advantages and disadvantages under different situations. On the basis of comprehensive analysis about techniques and methods used by the existing systems future development is discussed including the transformation of traditional recommendation, more refined modeling and personalization, and visualization and online computing.

REFERENCES.

[1] Zheng, X., Liang, X., & Xu, K. (2012, August). Where to Wait for a Taxi?. InProceedings of the ACM SIGKDD International Workshop on Urban Computing (pp. 149-156). ACM.

[2] Li, B., Zhang, D., Sun, L., Chen, C., Li, S., Qi, G., & Yang, Q. (2011, March). Hunting or waiting? Discovering passenger-finding strategies from a large-scale real-world taxi dataset. In Pervasive Computing and Communications Workshops (PERCOM Workshops), 2011 IEEE International Conference on (pp. 63-68). IEEE.

[3] Veloso, M., Phithakkitnukoon, S., Bento, C., Fonseca, N., & Olivier, P. (2011). Exploratory study of urban flow using taxi traces. In First Workshop on Pervasive Urban Applications (PURBA) in conjunction with Pervasive Computing, San Francisco, California, USA.

[4] Ding, Y., Liu, S., Pu, J., & Ni, L. M. (2013, June). Hunts: A trajectory recommendation system for effective and efficient hunting of taxi passengers. In Mobile Data Management (MDM), 2013 IEEE 14th International Conference on (Vol. 1, pp. 107-116). IEEE

[5] Yuan, J., Zheng, Y., Zhang, C., Xie, W., Xie, X., Sun, G., & Huang, Y. (2010, November). T-drive: driving directions based on taxi trajectories. InProceedings of the 18th SIGSPATIAL International conference on advances in geographic information systems (pp. 99-108). ACM.

[6] Powell, J. W., Huang, Y., Bastani, F., & Ji, M. (2011). Towards reducing taxicab cruising time using spatio-temporal profitability maps. In Advances in Spatial and Temporal Databases (pp. 242-260). Springer Berlin Heidelberg.

[7] Li, X., Pan, G., Wu, Z., Qi, G., Li, S., Zhang, D., ... & Wang, Z. (2012). Prediction of urban human mobility using large-scale taxi traces and its applications. Frontiers of Computer Science, 6(1), 111-121.

[8] Yuan, J., Zheng, Y., Xie, X., & Sun, G. (2011, August). Driving with knowledge from the physical world. In Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining(pp. 316-324). ACM.

[9] Yuan, N. J., Zheng, Y., Zhang, L., & Xie, X. (2013). T-Finder: A recommender system for finding passengers and vacant taxis. Knowledge and Data Engineering, IEEE Transactions on, 25(10), 2390-2403.

[10] Ge, Y., Xiong, H., Tuzhilin, A., Xiao, K., Gruteser, M., & Pazzani, M. (2010, July). An energy-efficient mobile recommender system. In Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 899-908). ACM.

[11] Qu, M., Zhu, H., Liu, J., Liu, G., & Xiong, H. (2014, August). A cost-effective recommender system for taxi drivers. In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 45-54). ACM.

[12] Yuan, J., Zheng, Y., Zhang, L., Xie, X., & Sun, G. (2011, September). Where to find my next passenger. In Proceedings of the 13th international conference on Ubiquitous computing (pp. 109-118). ACM.

[13] Guo, D., Zhu, X., Jin, H., Gao, P., & Andris, C. (2012). Discovering Spatial Patterns in Origin ‐ Destination Mobility Data. Transactions in GIS, 16(3), 411-429.

[14] Chang, H. W., Tai, Y. C., & Hsu, J. Y. J. (2010). Context-aware taxi demand hotspots prediction. International Journal of Business Intelligence and Data Mining, 5(1), 3-18.

[15] Phithakkitnukoon, S., Veloso, M., Bento, C., Biderman, A., & Ratti, C. (2010). Taxi-aware map: Identifying and predicting vacant taxis in the city. InAmbient Intelligence (pp. 86-95). Springer Berlin Heidelberg.

[16] Castro, P. S., Zhang, D., & Li, S. (2012). Urban traffic modelling and prediction using large scale taxi GPS traces. In Pervasive Computing (pp. 57-72). Springer Berlin Heidelberg.

[17] Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. Knowledge and Data Engineering, IEEE Transactions on, 17(6), 734-749.

[18] Kantor, P. B., Rokach, L., Ricci, F., & Shapira, B. (2011). Recommender systems handbook. Springer.

[19] Su, X., & Khoshgoftaar, T. M. (2009). A survey of collaborative filtering techniques. Advances in artificial intelligence, 2009, 4.

[20] Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. Computer, 42(8), 30-37.

[21] Koren, Y. (2010). Collaborative filtering with temporal dynamics.Communications of the ACM, 53(4), 89-97.

[22] Liu, L., Andris, C., Biderman, A., & Ratti, C. (2009). Uncovering taxi driver's mobility intelligence through his trace. IEEE Pervasive Computing, 160, 1-17.

[23] Liu, L., Andris, C., & Ratti, C. (2010). Uncovering cabdrivers' behavior patterns from their digital traces. Computers, Environment and Urban Systems, 34(6), 541-548.

[24] Veloso, M., Phithakkitnukoon, S., & Bento, C. (2011, September). Urban mobility study using taxi traces. In Proceedings of the 2011 international workshop on Trajectory data mining and analysis (pp. 23-30). ACM.

[25] Ge, Y., Liu, C., Xiong, H., & Chen, J. (2011, August). A taxi business intelligence system. In Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 735-738). ACM.

[26] Lou, Y., Zhang, C., Zheng, Y., Xie, X., Wang, W., & Huang, Y. (2009, November). Map-matching for low-sampling-rate GPS trajectories. InProceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (pp. 352-361). ACM.

[27] Yuan, J., Zheng, Y., Zhang, C., Xie, X., & Sun, G. Z. (2010, May). An interactive-voting based map matching algorithm. In Proceedings of the 2010 Eleventh International Conference on Mobile Data Management (pp. 43-52). IEEE Computer Society.

[28] Zheng, K., Zheng, Y., Xie, X., & Zhou, X. (2012, April). Reducing uncertainty of low-sampling-rate trajectories. In Data Engineering (ICDE), 2012 IEEE 28th International Conference on (pp. 1144-1155). IEEE.

[29] Zheng, Y., Liu, Y., Yuan, J., & Xie, X. (2011, September). Urban computing with taxicabs. In Proceedings of the 13th international conference on Ubiquitous computing (pp. 89-98). ACM.

[30] Bao, J., Zheng, Y., Wilkie, D., & Mokbel, M. F. (2014). A survey on recommendations in location-based social networks. Submitted to GeoInformatica.

[31] Chen, C., Zhang, D., Zhou, Z. H., Li, N., Atmaca, T., & Li, S. (2013, March). B-Planner: Night bus route planning using large-scale taxi GPS traces. InPervasive Computing and Communications (PerCom), 2013 IEEE International Conference on (pp. 225-233). IEEE.

[32] Ziebart, B. D., Maas, A. L., Dey, A. K., & Bagnell, J. A. (2008, September). Navigate like a cabbie: Probabilistic reasoning from observed context-aware behavior. In Proceedings of the 10th international conference on Ubiquitous computing (pp. 322-331). ACM.