Prediction of urban human mobility using large-scale taxi traces and its applications

P.Sandeep Reddy, CH Venkata Navi,N Mahesh Babu
Associate Professor, Assistant Professor[2,3]
Dept. of CSE,
mail-id:sandeepreddycse@anurag.ac.in, chejarla.venkatanavi5@gmail.com, mahesh.nallagatla@gmail.com
Anurag Engineering College,Anatagiri(V&M),Suryapet(Dt),Telangana-508206

Abstract ; This paper investigates human mobility patterns in an urban taxi transportation system. This work focuses on predicting human mobility from discovering patterns of in the number of passenger pick-ups quantity (PUQ) from urban hotspots. This paper proposes an improved ARIMA based prediction method to forecast the spatial-temporal variation of passengers in a hotspot. Evaluation with a large-scale real- world data set of 4 000 taxis' GPS traces over one year shows a prediction error of only 5.8%. We also explore the applica- tion of the prediction approach to help drivers find their next passengers. The simulation results using historical real-worlddata demonstrate that, with our guidance, drivers can reducethe time taken and distance travelled, to find their next pas- senger, by 37.1% and 6.4%, respectively.

Keywords; urban traffic, GPS traces, hotspots, human mo- bility prediction, auto-regressive integrated moving average (ARIMA) and wireless networking technologies, a growing number ofcomputing devices and sensors are embedded in our dailyenvironments, and becoming ubiquitous. As a result, muchinformation regarding human mobility, such as location, mo- tion, and behaviors of vehicles, is becoming easily accessi- ble. From these digital footprints, it is feasible for researchers to extract social and community intelligence [1], rangingfrom urban environment dynamics [2,3] to social events [4,5].The use of taxis conveys much information about human urban mobility. Their movement traces can be easily obtainedfrom equipped GPS devices. For example, many taxi compa- nies in China are required to install a GPS device in each of their own taxis for administrative purposes. This provides aninfrastructure to record the current and historical taxi traces
data for predicting urban human mobility.

This paper investigates human mobility patterns in an ur- ban taxi transportation system. We focus on discovering pat-terns of pick-up quantity (PUQ) for those urban hotspots with

a relatively large number of passengers getting in or out of

Introduction

Smart city, an emerging worldwide technology, aims to pro-mote sustainable economic development and high quality of life through intelligent management of resources, where understanding human mobility is one of the most important aspects. With the rapid development of embedded systems,

taxis. We propose an adaptive watershed algorithm to cluster hotspots. This algorithm can naturally determine the edges of hotspots according to the variation of PUQ within an ur- ban area. Prediction of urban human mobility can not only help people to experience a comfortable travel, but also helpthe government to improve the planning of the transportationsystem in a city. We develop an improved auto-regressive in-

to forecast how many passengers will be in a certain hotspotin the next time interval. We also explore the application of

the prediction approach to helping drivers to find their next passengers.

Section 2 outlines related work. Section 3 describes the data set used in this paper, and introduces our data set prepro-cessing and hotspot extraction methods. Section 4 presents and evaluates the improved ARIMA based method for pre- dicting human mobility in a hotspot. An application which aims to help taxi drivers find next passengers is proposed andevaluated in Section 5. Finally we conclude the paper in Sec-tion 6.

al. [16] discovered anomalous driving patterns from taxi GPStraces, targeting applications such as automatic detection of taxi fraud or road network changes in modern cites.

There also exits work that aims to provide guidance to passengers or taxi drivers to make their life or work more convenient. Phithakkitnukoon et al. [17] focused on predict-ing the distribution of vacant taxis in the city with a naive Bayesian classifier, which considers several factors, such as weather, day of the week, and time of day, to improve per- formance. The prediction is performed on a large region par-

tition (1 km × 1 km). Chang et al. [18] predicted taxi de-

ISSN- 2394-5125 VOL 07, ISSUE 19, 2020

Related work

Since a large amount of human position data has become ac- cessible, the patterns of human movement have been investi- gated in recent years. Several approaches used mobile phone traces [6] to analyze human mobility pattern. With the finger- print of cell-phone, Girardin et al. [7] focused on the pattern of tourists present in a public place. González et al. [8] uncov- ered the spatial-temporal regularity of human mobility. Mc- Namara et al. [9] analyzed historical collocation information of people in a day and made media sharing more efficient with this information. Most of this work focuses on mining the internal structure of human movement.

Some work explores large scale public transportation data sets for analyzing the urban traffic environment. Froehlich et al. [10] investigated the dynamics of the city of Barcelona from a shared bicycling system. They analyzed the relation- ship between the behavior patterns and the location of bicy- cle stations, where four prediction models were used to fore- cast the number of available bicycles in stations. Based on the same bicycling system, Kaltenbrunner et al. [11] detected temporal and geographic mobility patterns within Barcelona and used an ARMA model to predict the number of bicycles in a station to help improve the spatial deployment of stations. Similar to mobile phone IDs, taxi GPS trajectories con- vey much useful information. Ziebart et al. [12] built valuable navigation services by reasoning on driver behavior. Yuan et al. [13] provided navigation services by extracting the expe- rience of taxi drivers from historical taxi GPS footprints. Liu et al. [14] revealed the strategy of taxi drivers by comparing the performance of top drivers and normal drivers. Zheng et al. [15] detected flawed urban planning using the GPS tra- jectories of taxicabs traveling in urban areas. These GPS trajectories can evaluate the effectiveness of urban planning, such as a newly built roads and subway lines in a city, and remind city planners of a problem that had not been previ- ously recognized when they conceive future plans. Zhang et

mand in urban environments. First, they filter the historical data set using current contexts, such as location, time, and weather. Then the filtered data are clustered and mapped to road names semantically. However the authors of [18] do not consider the distribution of vacant taxis around the clusters they provide, which influence the real demand. Ge et al. [19] presented a method to recommend a sequence of pick up points or potential parking positions to taxi drivers. By track- ing this sequence, taxi drivers pay the least expected cost to find their next passenger. Yuan et al. [20] presented a rec- ommender system, for taxi drivers and passengers wishing to hail a taxi, using the knowledge of passenger mobility pat- terns and taxi driver pick-up behaviors learned from the GPS trajectories of taxicabs. For a taxi driver, they recommend a parking place using a probability model which maximizes the profit of taxi drivers who takes the recommendation.

For the prediction problem, the work in [10,11,17] is sim- ilar to ours. We predict the pick-up/set-down rate of passen- gers at a hotspot, while [10] and [11] predict the number of bi- cycles in bicycle stations, and [17] predicts the number of va- cant taxis. According to the methods used in [18–20], the im- mediate historical data cannot be exploited to make a recom- mendation; however, our recommendation is mainly based on the most immediate historical data. In this way, our method can make a more timely reaction to abnormal changes in taxi and human mobility patterns.

Data set preprocessing and hotspotextraction

The taxi GPS trace data set used in this paper is providedby the Hangzhou City Traffic Bureau. Hangzhou, is lo- cated in he Southeast of China, and is the capital of Zhe- jiang Province. It is one of the most famous tourist cities in China. According to the annual report of the Hangzhou gov-ernment, there were more than 53.24 million tourists visit- ing Hangzhou from all over the world in 2009. The taxi GPS traces were generated over a period of 385 days (from April 1, 2009 to April 20, 2010) [21]. During this period, the number of taxis with GPS devices installed increased from 4 597 to 7 475, while the total number of taxis in the city remained al-most unchanged. The status of GPS enabled taxis is sampled

sliding window.

Incorrect records would result in a non-smooth trajec- tory with abnormal movement. Given a trajectory, Function f ($record_i$, $record_j$) is defined as

with a fixed time interval of approximately 60 s. In addition,

-1, $Velocity_{i,j} > Threshold, i \neq j$;

1, $Velocity_{i,j}$ “ $Threshold, i \neq j$; $_i$

set contains approximately three billion records. Each record

Data set preprocessing

Due to the multipath effect of GPS signal and device faults, the GPS position may sometimes be incorrect. In addition, the METER STATE may also be incorrect due to invalid oper- ations made by a driver. For example, when a taxi driver goesoff work, they may keep the taximeter turned on although there is no passenger in the taxi. To clarify the real vacant and occupied trajectories (trajectories with and without pas- senger, respectively), a data preprocessing step is performed as follows:

Step 1 Extract the raw taxi trajectories from GPS records.

ISSN- 2394-5125 VOL 07, ISSUE 19, 2020

A shift of METER STATE is taken as a pick-up/set-down event, i.e., a METER STATE change from 0 to 1 suggests a pick-up event, and a shift from 1 to 0 of METER STATE in-dicates a set-down event. An occupied trajectory is defined as a series of records beginning with a pick-up event and end- ing with a set-down event, otherwise a vacant trajectory is de-fined as a series of records from a set-down event to a pick-up event, illustrated in Eq. (1).

where $Velocity_{i,j} = L_1(Pos_i, Pos_j)\ (Time_i - Time_j)$, and$L_1(Pos_i, Pos_j)$ is the city block distance. *Threshold* is setto 120 km/h according to urban traffic regulation. Records with velocity greater than the threshold indicate that there is an abnormal movement, therefore they should be re- moved. We define W as the abnormal velocity indicator for the ith record, which is defined as the sum of the neighboring$f(record_i, record_j)$

$$w_i = \sum_{j=-r}^{j=r} f(record_i, record_j), \qquad (3)$$

where r is the width of the sliding window. If $w_i < 0$, $record_i$is abnormal and should be removed. We heuristically set r to3.

Step 3 Filtering taxis with high flipping.

A limited number of taxis may produce a dramatically large number of trajectories in a single day, which indi- cates that the taximeter flips at an abnormally high fre- quency. These indicate abnormal taxi operation, likely a faultwith the meter. These abnormal traces are removed.

Step 4 Filtering invalid occupied trajectories.

We filter the occupied trajectories whose duration and av- erage speed is out of a normal range. We analyzed the distri- bution of the duration and average speed of occupied trajec- tories. We found that a majority of all valid trajectories fall between 100–5 000s, so we use these as the limits for valid trajectories. Using similar analysis the average speed range is set to be 1.5 m/s to 40 m/s. All the occupied trajectories that do not satisfy these conditions will be considered to be invalid.

Hotspot extraction

Hotspots are urban areas in which pick-up/set-down events occur more frequently. The activities in hotspots can char- acterize the spatial mobility pattern of the whole city. Our

Step 2 Filtering incorrect records from a trajectory with a hotspot extraction procedure is asfollows:

1) 4 000 taxis are randomly sampled for analysis to avoidthe influence of variation in the quantity of taxis.
2) Pick-up/set-down events are extracted from historical taxi trajectories.
3) The map of Hangzhou is divided into blocks of 10 m × 10 m, pick-up and set-down events are tallied in PUQ and SDQ, respectively. The PUQ and SDQ for each block in a specified period are counted and blocks with a PUQ or SDQ greater than a threshold are labeled valid blocks. Adjacent valid blocks are then merged intorough hotspots.
4) An adaptive watershed algorithm, described in the next subsection, is employed to split rough hotspots into smaller hotspots. Some regions in an urban area havea relatively high value of PUQ/SDQ. If we set a sin- gle threshold for valid blocks, we will probably get very large rough hotspots in these regions. Taking roughhotspots directly as hotspots is not reasonable, becausethe value of PUQ/SDQ could vary greatly within the rough hotspots, which means they could contain sepa- rate hotspots.

Figure 1 shows some hotspots extracted from the area around the West Lake. The gray regions indicate the ex- tracted hotspots. Adjacent hotspots are plotted in different grayscales.

ISSN- 2394-5125 VOL 07, ISSUE 19, 2020

Fig. 1 Illustration of hotspots extracted from around the West Lake (hotspots are marked in gray). Adjacent hotspots are plotted in different grayscales

3.2.1 Adaptive watershed algorithm for hotspot splitting

The watershed algorithm is a traditional image processing technique to segment an image [22]. In image segmentation, a gray-level image is regarded as a topography, the gray levelof a pixel indicates its altitude; when flooding this supposi- tional topography from the bottom, the edge of this image canbe sketched by building a barrier to avoid the connection of adjacent basins, since the edges of a gray-level image have the local maximum gray level.

To obtain hotspots from rough hotspots, a rough hotspot with PUQ/SDQ density can be regarded as a topogra- phy. Splitting of the rough hotspot is to determine the edges of this suppositional topography. Different from the edges ofa gray-level image, the edges in rough hotspots are a set of blocks whose PUQ/SDQ is locally minimal. To avoid over- splitting, we set a minimal radius for a rough hotspot; any rough hotspot whose radius is smaller than the threshold willnot be split.

Algorithm 1: Adaptive watershed-based hotspot splittingalgorithm

nput: *B* is the set of blocks in rough hotspot *C*

```
Hotspots ← Φ;
Blocks ← sort(B.PUQ, descend);
Unlabeled ← B;
Labeled ← Φ;
while is_connected(Unlabeled) /*test if blocks in Unla-beled are 8-connected*/
    b ← min(Unlabeled.PUQ);
    Labeled. join(b);
    Unlabeled.remove(b);
S ← subsets(Unlabeled);   /*find out subsets of blocks inwhich blocks are 8-connected*/
n ← Blocks.size;
if !is_empty(Labeled)
   for i = 1 to n do
      if !is_labeled(Blocks(i))
         neighbors ← get_neighbors(Blocks(i for all nei ∈ neighbors
            if is_labeled(nei)
               subset_index ← get_subset_index
               (Blocks(i));
               S (subset_index).join(nei);
               Labeled.remove(nei);
               Unlabeled. join(nei);
for all s ∈ S
   if s.radius > minimal
      Blocks ← s;
      goto 2;
   else
```

```
        Hotspots. join(s);
    Return Hotspots;
```

Given a rough hotspot, this hotspot contains a set of connected blocks; Algorithm 1 illustrates the adaptive watershed-based hotspot splitting algorithm.

Prediction of human mobility for a hotspot

To measure the temporal spatial human mobility in Hangzhou, a day of 24 hours is uniformly divided into *D* timesegments (TS) with a *D*-hour length, where *D* is called the time segment length. For example, when *D* is 3 h, a day is divided into eight time segments; when *D* is 20 min, one dayis divided into 72 time segments.

In this section, four algorithms are evaluated to predict the PUQ value of hotspots. We use a naive method, Bayesian net- works, auto-regressive integrated moving average (ARIMA)[23], and an improved ARIMA. ARIMA is a classical ap- proach for time series analysis, and the improved ARIMA is an improvement over ARIMA by considering the repeatedpattern of PUQ in hotspots.

Prediction methods

Naive method

Figure 2 depicts the mean and variance of PUQ and SDQof two hotspots throughout a day. The relatively low vari- ance indicates that the PUQ of a hotspot has a repetitive pattern with the period of one day. Given a time series of PUQ $\{PUQ_i, i = 1, 2, \ldots, N\}$, a straightforward way to pre- dict PUQ at a future time segment is to use the PUQ one daybefore that time segment.

$$P\hat{U}Q_{N+1} = PUQ_{N+1-(24/D)}. \qquad (4)$$

Here, *D* is time segment length in hour. For example, sup- pose that the time segment length is set to 1 h, using the naive method the predicted value of PUQ at the time segment *n* is equal to the value of PUQ in the time segment *n* – 24, i.e., atthat time yesterday.

Bayesian networks

Bayesian networks are widely used to represent the relation- ships between random variables. They can be used for pre- dictions in time series analysis [10]. Given a time series of PUQ $\{P\hat{U}Q_i, i = 1, 2, \ldots, N\}$, PUQ_{i+1} can be forecasted by the single layer Bayesian networks model in Fig. 3, in which$CT_{i+1} = (i + 1) \bmod (24/D)$ is the index of time segment *i*+1 in a day.

Original ARIMA based prediction

ARIMA is widely used in time series analysis [23]. It is an in-tegrated generalization of the auto-

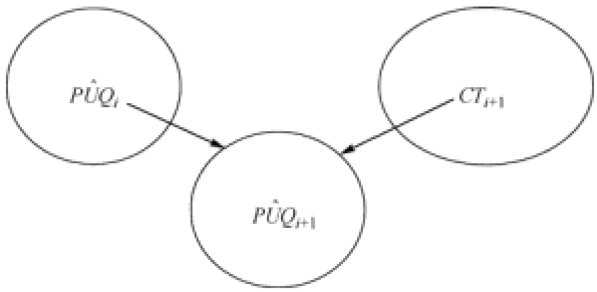

Fig. 3 Bayesian networks for prediction of PUQ

the prediction of $P\hat{U}Q_N$. The ARIMA algorithm [23] de- scribed as below:

$$\varphi(B)\nabla^d P\hat{U}Q_i = \theta(B)\alpha_i, \qquad (5)$$

where B is the lag operator, $\varphi(B)$ is the auto-regressive pro-cess, ∇^d is the differencing operator, $\theta(B)$ is moving average

where $P\hat{U}Q_{d,t+1}$ is the predicted value of $PUQ_{d,t+1}$, $\hat{q}_{d,t+1}$ is the predicted value of $q_{d,t+1}$, and $CDF_t^{-1}(x)$ is the inverse function of $CDF_t(x)$.

The number of pick-ups in a hotspot H is influenced by many factors, such as the social function of the area surround-ing H, time of day, weather conditions, weekday or holiday, and special events. We can roughly group these factors into two categories: major factors and secondary factors. Major factors are the social functions around H and the time of day, secondary factors are all the other factors, some of which can-not be observed.

$$PUQ_{d,t} = f_m(L, t) + f_s(\boldsymbol{V}), \qquad (10)$$

where L is the social functions around H, t is the time of day,and $\boldsymbol{V}$ is the vector of secondary factors. $f_m(L, t)$ is a periodic process, and α_i is a random walk process [23]. Eq. (5) de-

scribes the relation between future and historical values of $P\hat{U}Q_i$. According to Eq. (5), the predicted value $P\hat{U}Q_N$ can be obtained from historical values of PUQ_N.

Our improved ARIMA based method

Considering the periodicity of the number of pick-up events at a hotspot, PUQ_i is denoted as

$$PUQ_{d,t}, \quad d = 1, 2, \ldots, n; \quad t = 1, 2, \ldots, m, \qquad (6)$$

$$i = (d - 1) \times n + t, \qquad (7)$$

where d is the index of the day, t is the index of time segmentin a day. The problem is to forecast the value of $PUQ_{n,t+1}$. The prediction of ARIMA is based on the value of PUQ in the nearest past few time segments. Notice the periodicity of PUQ/SDQ illustrated in Section 4, we improve ARIMA by considering not only the nearest historical data but also the periodicity of PUQ/SDQ. The improved ARIMA method is described as follows:

- For each time segment t, the cumulative distribution function (CDF) is extracted by non-parametric estima-tion from the past $n - 1$ days. And the value of the CDF at $PUQ_{d,t}$ is calculated.

$$q_{d,t} = CDF_t(PUQ_{d,t}) = P(X_t < PUQ_{d,t}), \qquad (8)$$

where $CDF_t(x)$ is the CDF of PUQ in time segment t over the past $n - 1$ days.

- Forecast the value of $q_{n,t+1}$ with the original ARIMA.
- Obtain the predicted value of $PUQ_{n,t+1}$ from $q_{n,t+1}$ with $CDF_t^{-1}(x)$

ISSN- 2394-5125 VOL 07, ISSUE 19, 2020

$$P\hat{U}Q_{d,t+1} = CDF_t^{-1}(\hat{q}_{d,t+1}), \quad (9)$$

function, and the time period is 24 hours.In Step 1,

$$\begin{aligned} q_{d,t} &= CDF_t(PUQ_{d,t}) \\ &= P(X_t < PUQ_{d,t}) \\ &= P(X_t < f_m(L, t) + f_s(\boldsymbol{V})) \\ &= P((X_t - f_m(L, t)) < f_s(\boldsymbol{V})) \\ &= P(X_t^J < f_s(\boldsymbol{V})), \end{aligned} \quad (11)$$

where $X_t^J = X_t - f_m(L, t)$, X_t is the random number of pick-upsin time segment t of a day. Due to the periodicity of $f_m(L, t)$, with regard to a specific time segment t, $f_m(L, t)$ is a constant, so X^J is a random variable of $f_s(\boldsymbol{V})$, which means $q_{d,t}$ is a function of $\boldsymbol{V}$ and can be denoted as

$$q_{d,t} = \varphi(\boldsymbol{V}). \quad (12)$$

Equation (12) illustrates that $q_{d,t}$ measures the effect of $\boldsymbol{V}$ sep-arate from the influence of time segment and social functionsof H.

Evaluation

Evaluation methodology

We selected 100 extracted hotspots with high PUQ for pre- diction evaluation. Two error measurements are employed to evaluate the prediction accuracy.

(1) Symmetric mean absolute percentage error ($sMAPE$) [24],which is defined as

$$sMAPE = \frac{1}{n}\sum_{i=1}^{n} \frac{PUQ - PUQ}{PUQ_i + P\hat{U}Q_i}$$

We set the training set size of the Bayesian network to be 350 days, the training length of the ARIMA and our improved ARIMA methods to be 4, 7, and 14 days, the time segment length to be 1, 2, 3, 6, 12, and 24 hours.

Performance comparison

Figure 4 compares the performance of the four methods us- ing two error measures, $sMAPE$ and $NMAE$. Our improved ARIMA achieves the best performance of the four methods both in $sMAPE$ and $NMAE$. The prediction of ARIMA is based on the PUQ value of recent historical data, while the improved ARIMA based prediction method considers not only the recent historical data but also the PUQ periodic- ity. Due to the obvious periodicity of PUQ, the naive methodalso achieves good performance especially when the time segment is long. A shorter time segment usually leads to a more random value of PUQ. As a result, when the length of time segment increases, PUQ suggests a more regular period-icity, which gives the naive method a good performance. Thepoor performance of Bayesian networks partially results fromits simple structure. However, if we consider more factors in

An application: helping taxi drivers findnext passengers

It is important for taxi drivers to find their next passenger assoon as possible, since they can make more money in less time using less fuel. Skilled taxi drivers know roughly where and when there will likely be passengers nearby, since they are familiar with the mobility pattern of passengers in an

urban area. However, inexperienced drivers do not have suchknowledge. Our prediction approach can be exploited to helptaxi drivers to find their next passengers more effectively bypredicting the PUQ of those hotspots near the current taxi po-sition. In our application, we use a waiting strategy: driving to the suggested hotspot and waiting there to pick-up pas- sengers. Skilled drivers can also benefit from our approach, because the mobility pattern is dynamic and varies over time.

Problem definition

Assume that $\{h_i, i = 1, 2, \ldots, m\}$ is a set of hotspots, $\{p_i, i = 1, 2, \ldots, m\}$ is the position of h_i, $\{P\hat{U}Q_{t,i}, i = 1, 2, \ldots, m\}$ is the predicted value of PUQ in h_i in time segment t. Given a

speed from the block containing P to the block containing h_i;if there are no historical trajectories from one block to an- other, the speed is replaced with the average speed between the nearest blocks.

- Compute waiting time

There are two main factors influencing the expected wait-ing time; they are the PUQ and the length of the waiting queue of vacant taxis in the hotspot. Since we assume that taxis join a FIFO queue at a new hotspot, the customer will take the first taxi in the queue. According to the assumptions both taxis and passengers arrive subject to a negative expo- nential distribution.

Given a hotspot h_i and time segment t, suppose that

context (t_0, P), where the current time t_0 is in time segment t,and P is a position, which means that a taxi

to the result of guiding taxis with the predicted result of ARIMA, the result with improved ARIMA achieves clear im- provement in the small hours. The main reason is that im- proved ARIMA outperforms original ARIMA in small hoursin predicting PUQ in hotspots (see Fig. 7).

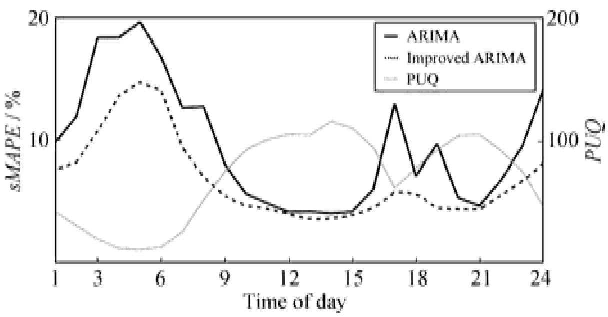

Fig. 7 Variation of *sMAPE* prediction error over a day

Conclusions

This paper addresses the prediction and application of hu- man mobility using large-scale taxi GPS traces. An adap- tive watershed-based hotspot extraction algorithm is pro- posed to cluster the pick-up/set-down events of taxi passen- gers. Four prediction methods, naive method, Bayesian net- works, ARIMA, and our improved ARIMA, are used to pre-dict the pick-up quantity of taxi passengers for hotspots. The improved ARIMA combines ARIMA with a prior distribu- tion of pick-up values, and achieves better prediction accu- racy than the other three methods.

Based on the prediction method, an application of helpingtaxi drivers find the next passenger is presented and evalu- ated. The evaluation using historical taxi GPS traces suggeststhat the time cost for finding passengers can be decreased by 37.1% and the length of vacant driving distance decreased by 6.4%. Actually, the prediction of hotspot passengers is notonly helpful for drivers, but also for traffic police, passengers,and even urban planning. Our future work plans to consider the possibility of picking up passengers on the way to the suggested hotspot.

References

Ratti C, Pulselli R M, Williams S, Frenchman D. Mobile Landscapes: using location data from cell phones for urban analysis. Environment and Planning B: Planning and Design, 2006, 33(5): 727–748

Zhu H, Zhu Y, Li M, Ni L. SEER: metropolitan-scale traffic perception based on lossy sensory data. In: Proceedings of the 28th Conference on Computer Communications. 2009, 217–225

Calabrese F, Pereira F C, Lorenzo G D, Liu L, Ratti C. The geography of taste: analyzing cell-phone mobility and social. In: Proceedings of the 8th International Conference on Pervasive Computing. 2010, 22–37

Girardin F, Blat J, Calabrese F, Fiore F, Ratti C. Digital Footprinting: uncovering tourists with user-generated content. IEEE Pervasive Com-puting, 2008, 7(4): 36–43

Ahas R, Aasa A, Silm S, Tiru M. Mobile positioning data in tourism studies and monitoring: case study in Tartu, Estonia. In: Proceedings of International Conference on Information and Communication Tech-nologies in Tourism. 2007, 119–128
Girardin F, Vaccari A, Gerber A, Biderman A, Ratti C. Quantifying urban attractiveness from the distribution and density of digital foot- prints. International Journal of Spatial Data Infrastructures Research, 2009, 4: 175–200
González M, Hidalgo C, Barabasi A. Understanding individual humanmobility patterns. Nature, 2008, 453: 779–782
McNamara L, Mascolo C, Capra L. Media sharing based on colloca- tion prediction in urban transport. In: Proceedings of the 14th ACM Annual International Conference on Mobile Computing and Network-ing. 2008, 58–69
Froehlich J, Neumann J, Oliver N. Sensing and predicting the pulse ofthe city through shared bicycling. In: Proceedings of the 21st Interna- tional Joint Conference on Artificial Intelligence. 2009, 1420–1426
Kaltenbrunner A, Meza R, Grivolla J, Codina J, Banchs R. Urban cy- cles and mobility patterns: exploring and predicting trends in a bicycle-based public transport system. Pervasive and Mobile Computing, 2010,6(4): 455–466
Ziebart B, Maas A, Dey A, Bagnell J. Navigate like a cabbie: proba- bilistic reasoning from observed context-aware behavior. In: Proceed-ings of the 10th ACM International Conference on Ubiquitous Com- puting. 2008, 322–331
Yuan J, Zheng Y, Zhang C, Xie W, Xie X, Sun G, Huang Y. T-Drive: driving directions based on taxi trajectories. In: Proceedings of the 18thACM International Conference on Advances in Geographic Informa- tion Systems. 2010, 99–108
Liu L, Andris C, Ratti C. Uncovering cabdrivers' behavior patterns from their digital traces. Computers, Environment and Urban Systems,2010, 34(6): 541–548
Zheng Y, Liu Y, Yuan J, Xie X. Urban computing with taxicabs. In: Proceedings of the 13th ACM International Conference on UbiquitousComputing. 2011, 89–98
Zhang D, Li N, Zhou Z, Chen C, Sun L, Li S. iBAT: detecting anoma-lous taxi trajectories from GPS traces. In: Proceedings of the 13th ACM International Conference on Ubiquitous Computing. 2011, 99– 108
Phithakkitnukoon S, Veloso M, Bento C, Biderman A, Ratti C. Taxi- Aware Map: identifying and predicting vacant taxis in the city. In: Pro-ceedings of the 1st International Joint Conference on Ambient Intelli-gence. 2010, 86–95
18. Chang H W, Tai Y C, Hsu Y J. Context-aware taxi demand hotspots
1. Zhang D, Guo B, Yu Z. The emergence of social and community intel-ligence. Computer, 2011, 44(7): 21–28
prediction. International Journal of Business Intelligence and DataMining, 2010, 5(1): 3–18
Ge Y, Xiong H, Tuzhilin A, Xiao K, Gruteser M. An energy-efficientmobile recommender system. In: Proceedings of the 16th ACM Inter-

national Conference on Knowledge Discovery and Data mining. 2010,899–908
Yuan J, Zheng Y, Zhang L, Xie X, Sun G. Where to find my next pas- senger? In: Proceedings of the 13th ACM International Conference onUbiquitous Computing. 2011, 109–118
Qi G, Li X, Li S, Pan G, Zhang D. Measuring social functions of city regions from large-scale taxi behaviors. In: Proceedings of the 9thIEEE International Conference on Pervasive Computing and Commu-nications, WiP. 2011, 384–388
Beucher S, Lantuejoul C. Use of watersheds in contour detection. In: Proceedings of the International Workshop on Image Processing: Real-time Edge and Motion Detection/Estimation. 1979
Box G, Jenkins G, Reinsel G. Time Series Analysis: Forecasting and Control. 4th ed. Hoboken: John Wiley & Sons, 2008
Makridakis S, Hibon M. The M3-Competition: results, conclusions and implications. International Journal of Forecasting, 2000, 16(4): 451–476
Cooper R. Introduction to Queueing Theory. New York: Macmillan, 1972