Hunting or Waiting? Discovering Passenger-Finding Strategies from a Large-scale Real-world Taxi Dataset

Bin LI[1], Daqing ZHANG[1], Lin SUN[1], Chao CHEN[1], Shijian LI[2], Guande QI[2] and Qiang YANG[3]
[1] *Institut Telecome SudParis, Evry, France*
{bin.li, daqing.zhang, lin.sun}@it-sudparis.eu
[2] *Zhejiang University, Hangzhou, China*
[3] *Hong Kong University of Science and Technology, Hong Kong, China*

***Abstract*—In modern cities, more and more vehicles, such as taxis, have been equipped with GPS devices for localization and navigation. Gathering and analyzing these large-scale real-world digital traces have provided us an unprecedented opportunity to understand the city dynamics and reveal the hidden social and economic "realities". One innovative pervasive application is to provide correct driving strategies to taxi drivers according to time and location. In this paper, we aim to discover both efficient and inefficient passenger-finding strategies from a large-scale taxi GPS dataset, which was collected from 5350 taxis for one year in a large city of China. By representing the passenger-finding strategies in a Time-Location-Strategy feature triplet and constructing a train/test dataset containing both top- and ordinary-performance taxi features, we adopt a powerful feature selection tool, L1-Norm SVM, to select the most salient feature patterns determining the taxi performance. We find that the selected patterns can well interpret the empirical study results derived from raw data analysis and even reveal interesting hidden "facts". Moreover, the taxi performance predictor built on the selected features can achieve a prediction accuracy of 85.3% on a new test dataset, and it also outperforms the one based on all the features, which implies that the selected features are indeed the right indicators of the passenger-finding strategies.**

Keywords-Taxi Data Mining, GPS, Passenger-Finding Strategy, Large-scale Data, Reality Mining

I. INTRODUCTION

GPS has become a powerful ubiquitous sensor in our daily life. Nowadays, it has been embedded into many smart phones, vehicles, and other devices. GPS devices are playing a front-line role to continuously create and record the digital footprints of the carrier. Many daily facets of the carrier can be inferred from these spatio-temporal data. Since the most commonly used GPS devices by average population are smart phones and vehicles, which are extensively deployed in metropolitan areas, they can thus be a very rich data source for understanding city dynamics and revealing hidden social and economic "realities". Previous works along this line include similar place identification (Mobile Landscapes) [1], mobility pattern analysis [2-4], traffic condition prediction [5-9], and taxi-mobility intelligence [10-13].

Public transportation is one of the most popular and important application fields of GPS. In modern cities, many public transportation vehicles, such as taxis, have been equipped with GPS devices. In the beginning, GPS devices equipped in taxis were mainly used for localization, navigation, scheduling and planning. As more and more taxis are equipped with GPS sensors and wireless communication units, immense amount of taxi status and GPS trajectory data can be collected in real-time. Based on the collected GPS trajectory dataset, besides the applications like "location-based services" based on the real-time context of individual taxis, some new applications leveraging the mobility pattern of a large collection of taxi GPS data are emerging, ranging from urban design to traffic prediction. Many interesting research issues have been explored based on the large-scale taxi GPS trajectories, such as hotspots and traffic condition detection [7-9], and taxi mobility intelligence mining [10-13].

In this paper, we intend to investigate what are the efficient and inefficient passenger-finding strategies based on a large-scale real-world taxi GPS dataset, which was collected from 5350 taxis for one year in a large city of China. In particular, we would like to study what feature sets have most significant impact on taxi drivers' performance and how these feature sets can be used to guide the ordinary taxi drivers to improve their driving strategies, based on the pattern analysis of top- and ordinary-performance taxi group dataset. With this objective in mind, the big challenge is how to translate the underlying "strategy" into the machine-understanding formalism, such that it can be processed by the appropriate data mining algorithms to discover the latent knowledge. To this end, we first symbolize various passenger-finding strategies into a collection of feature patterns represented by triplet (Time, Location, Strategy). The rationale behind is to study what taxi drivers' behavior in a certain timeslot and location will lead to good or ordinary performance. We list all the combinations of the three attributes and count the times that a taxi falls in each pattern. We then adopt a powerful feature selection tool, L1-Norm SVM, to select the most salient patterns for discriminating top- and ordinary-performance taxis. We find that the selected patterns can well interpret the empirical study results derived from raw data analysis and even reveal interesting hidden "facts". Moreover, the taxi performance predictor built on the selected features can achieve a prediction accuracy of 85.3% on a new test dataset, and it also outperforms the one based on all the features, which implies that the selected features are indeed the right indicators of the passenger-finding strategies.

The remainder of the paper is organized as follows. In Section II, we will have a brief review on the related work. A large-scale real-world taxi data is introduced in Section III and an empirical study with visualization is presented in

Section IV. In Section V, we symbolize a collection of passenger-finding strategies and adopt a feature selection technique to select the most salient feature patterns for determining the performance of taxis, accompanied with analysis. We conclude this work in Section VI.

II. Related Work

Urban vehicle transportation and service understanding: Yamamoto et al. [14] proposed an adaptive routing algorithm using fuzzy clustering to improve taxi dispatching by assigning vacant taxis to pathways with many potential customers expected. *Chang et al.* [8] described a model that predicts taxi demand distribution based on time, weather condition, and location. *Santi et al* [9] used taxi data to identify and predict vacant taxis in the city. *Yang et al* [10,11] aimed to model urban taxi services in a network context which can describe vacant and occupied taxi movements in a road network and the relationship between customer and taxi waiting times.

Taxi Driver mobility intelligence: Few works have been reported on this topic. The closest work to ours is [12,13] for uncovering taxi drivers' mobility intelligence, which has the similar goal as ours. They use *K*-means clustering technique to analyze the spatio-temporal patterns of the GPS data and understand the operation patterns of taxi drivers [12]. They also use a GPS dataset for one year to reveal the behaviors of taxi drivers by analyzing continuous digital traces [13]. The differences between [12,13] and ours are: 1) We design a triplet descriptor to symbolize the pickup/dropoff events as the features, while they employ other features of trajectories. 2) We use feature selection technique to discover the most salient feature patterns for determining the performance of taxis, while they only predict the taxi performance based on the full set of features. 3) Last and most important, we model the subjective passenger-finding strategies and discover the useful ones which can be guidelines for taxi drivers, while they focus on revealing what characteristics top drivers have. Different from the previous papers, our work is the first one that investigates what features have most significant impact on taxi drivers' performance and how these features can be used to guide taxi drivers to improve their driving strategies.

III. Data Preprocessing & Extraction

We get a large-scale real-world taxi GPS dataset of more than 5350 taxis served in a large city in China (Hangzhou) for one year (Apr 2009 ~ Mar 2010). Hangzhou has a population of more than 6 million people and it is also a famous tourism city. The large population and massive passenger flows raised great challenges and opportunities to taxi drivers. Good or bad passenger-finding strategies can significantly influence taxi drivers' performances.

In our taxi dataset, each taxis is deployed with a GPS device for recording real-time taxi information at a sampling frequency of 1~7 times per minute. Each record contains the following fields which we will use in this research work:

- VEHICLE_ID: unique ID of the taxi in the dataset;
- LONGITUDE: current longitude of the taxi;
- LATITUDE: current latitude of the taxi;
- SPEED: current speed of the taxi;
- STATE: current status (occupied/vacant) of the taxi;
- SPEED_TIME: sampling timestamp in the format of "YYYY-MM-DD HH:MM:SS".

There are more than 200 billion records in the entire dataset. To reduce the computing burden, we choose the records of 15 working days in Oct 2009 as our research dataset. To limit our interested area only in Hangzhou metropolitan area, we choose the records with both pickup and dropoff locations within the area of longitude [120,120.5] and latitude [30.15,30.40]. The records out of the selected time and location range are discarded. We also discard the records which may be caused by device errors and noises. After data pruning, we obtain 4548 taxis with more than 500 valid GPS records in the selected 15 days.

In this paper, we are interested in studying the passenger-finding strategies, in particular, the taxi drivers' behaviors before picking up and after dropping off passengers. And we don't investigate the driving trajectories. Thus, we only need to extract the pickup and dropoff events for each ride.

For each pickup/dropoff event, we extract the GPS data and event timestamp. Besides, we also estimate the driving distance during 3 minutes before a pickup event and the driving distance during the period after current dropoff event and before the next pickup event. With these two distance indicators and proper thresholds (discussed in Section V), for pickup event, we can know whether a taxi is waiting at a location or roaming around, while for dropoff event, we can know whether a taxi is directly heading for somewhere for next passengers or just haunting near the dropoff location. The extracted information from the raw GPS records for pickup/dropoff events are represented in a table (Table 1), where tag value 1 denotes a pickup event and -1 a dropoff event. For example, the first row means that "at 00:19:48, 12 Oct 2009, taxi 20731 picked up a passenger at [120.107590, 30.320194], and 3 minutes before this pickup event, the taxi had run 0.68km". The second row means that "at 00:38:15, 12 Oct 2009, taxi 20731 dropped off a passenger at [120.191025, 30.265770], and from this dropoff event to the next pickup event, the taxi had run 1.99km" .

TABLE I. An example of pickup/dropoff records

Taxi	Longitude	Latitude	Year	Month	Day	Hour	Min	Sec	Tag	Dist
20731	120.107590	30.320194	2009	10	12	0	19	48	1	0.68
20731	120.191025	30.265770	2009	10	12	0	38	15	-1	1.99
20731	120.179450	30.268055	2009	10	12	0	49	53	1	0.11
20731	120.156630	30.256460	2009	10	12	0	58	5	-1	2.94

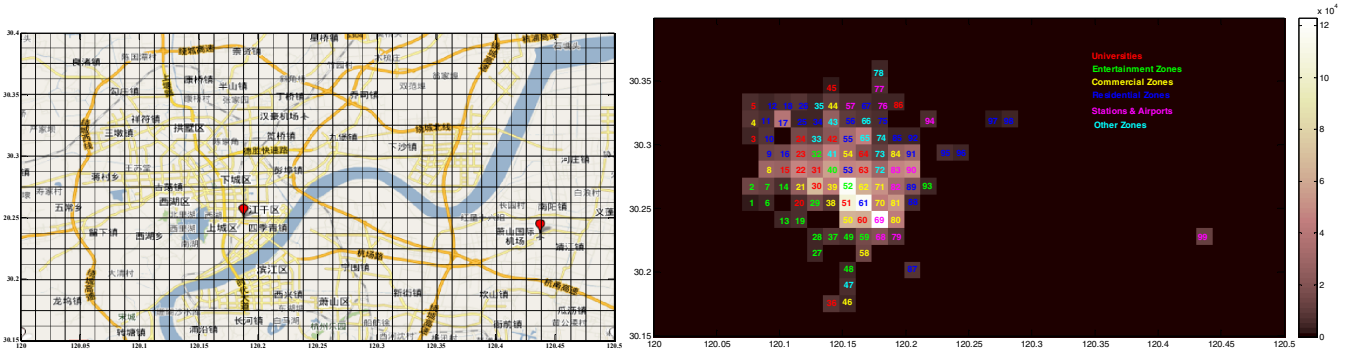

Figure 1. (Left) Hangzhou metropolitan area partition (40×20 grids). Two red spots are the raliway station (left) and the int'l airport (right), respectively. (Right) Density map of Hangzhou metropolitan area. The grayscales indicate the pickup/dropoff times in a certain region in the selected 15 days. We number the top 99 busiest regions and the rest non-hot area is treated as one region with label 100. Different label colors indicate different region functions. The railway station is in region 69 and the int'l airport is in region 99.

IV. EMPIRICAL STUDY

A. *Hotspots Analysis*

We partition Hangzhou metropolitan area (longitude [120.0,120.5], latitude [30.15,30.4]) into 40×20 grids with equal intervals (see Figure 1). We count all the pickup and dropoff events during the selected 15 days in each region and select the top 99 busiest regions as the hotspots and the rest as non-hot area. The locations and functions of the hotspots are numbered as shown in Figure 1. The far away isolated region 99 is the int'l airport. The railway station, commercial zones, residential zones, and the main campus of Zhejiang University surrounding the West Lake are the top hotspots.

The skeleton of Hangzhou metropolitan area can be outlined by all the pickup/dropoff points as shown in Figure 2. We can see that the dropoff points (blue) are more scattered than the pickup points (red) since people usually catch a taxi on main roads while get off anywhere. We count the pickup/dropoff points in each region at 6 equal time slots (sub-captions in Figure 2). We plot the top 10 pickup/dropoff hotspots on the background of pickup/dropoff point cloud in Figure 2. Most of the top 10 pickup/dropoff regions are surrounding the West Lake across different time slots. The top 10 pickup/dropoff hotspots also depend on the time. The railway station is among the top 10 pickup/dropoff hotspots across the whole day. During mid-night (00h~04h), as expected, the dropoff locations are almost located in residential zones while the pickup locations are mainly in the railway station and entertainment zones. During rush hours (04h~08h), the pickup locations are mainly in residential zones while the dropoff locations are mainly in commercial zones. As seen from Figure 2(b), the airport is among top 10 hotspots only during the early morning. The reason may be that passengers are inclined to take airport shuttle buses in daytime since it is cheaper. While during the early morning people turn to take taxis since airport shuttle buses are out of services.

B. *Hunting or Waiting?*

Two passenger-finding strategies, namely hunting and waiting, are mainly adopted by taxi drivers when they are trying to find passengers. By analyzing the average pickup numbers during one certain period of time in one certain location with respect to both strategies, we can get a clear view of which one is more effective in the context. We estimat this average value with our data in the following method. During a certain time period of each day in one certain region, we count the number of pickup events separately with respect to hunting and waiting in the selected 15 days. We also calculate the total waiting hours and hunting hours of these taxis. Then we divide the counted numbers with each waiting and hunting hours respectively and get the averaged pickup times per hour in terms of hunting and waiting. The comparison result is shown in Figure 3. In each subplot, the X-axis is the hotspot label defined in Figure 1. Those locations are the top 20 pickup density locations in the corresponding time slots. The Y-axis is the average passenger pickup times.

From Figure 3 we can see that during 22h~23h and 00~01h, the total pickup number is smaller than the other time periods and hunting performs better than waiting in most of the hotspots. This reflects that the good traffic condition in this time period may help the taxi drivers find passengers when they hunt and also as the passengers are fewer compared to those in rush hours, waiting in one location is unlikely to find passengers. During 06h~07h hunting performs far better than waiting since it's the time when people go to work from their scattered homes. One exception is the railway station. It is because the railway station has strict rules for taxis to wait in line for the coming passengers and those taxis' behaviors are defined as waiting since they hardly move in the 3 minutes before picking up passengers.

During 10h~11h, 12h~13h and 18h~19h time periods, waiting performs better than hunting for the hottest spots, and as the grids become less hot, hunting performance increases while waiting performance decreases. And finally hunting performs better than waiting. This may be because those hottest spots are normally associated with traffic congestion. As a result, the taxis are normally regarded as waiting since their moving speed is too slow. On the contrary, the traffic conditions of these less hot regions are normally good and taxi drivers can travel around to find passengers.

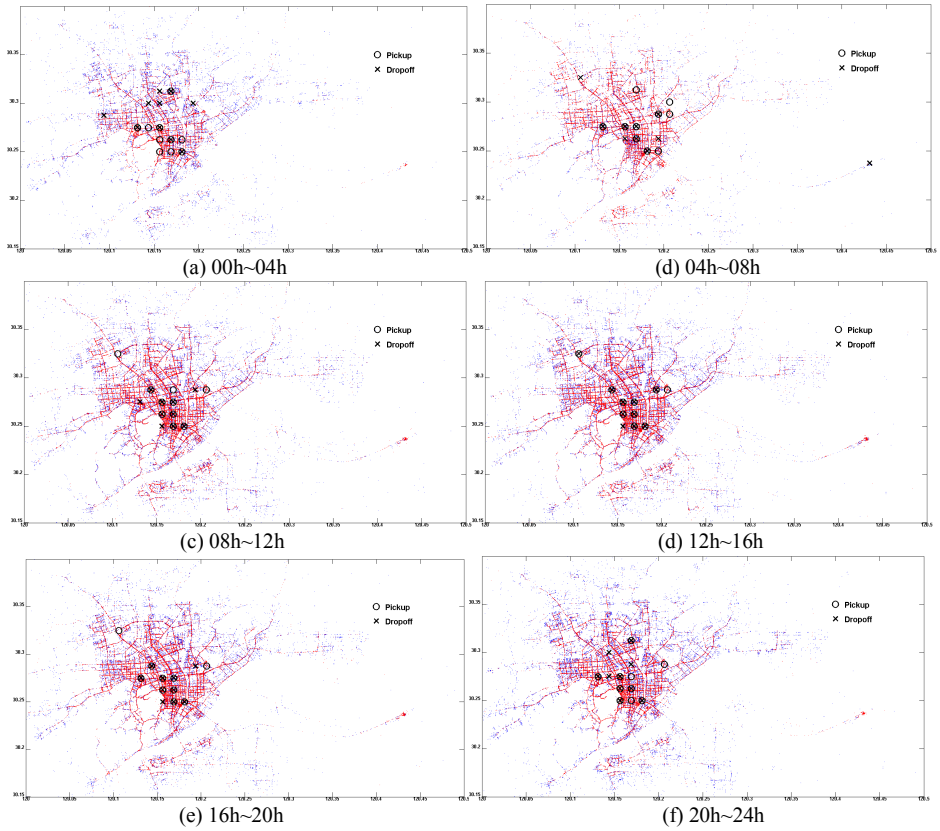

Figure 2. Top 10 pickup/dropoff hotspots (matked with 'o'/'x') at different time slots in Hangzhou metropolitan area. The background point clouds are plotted by using all the pickup points (red) and dropoff points (blue) at the same time slot.

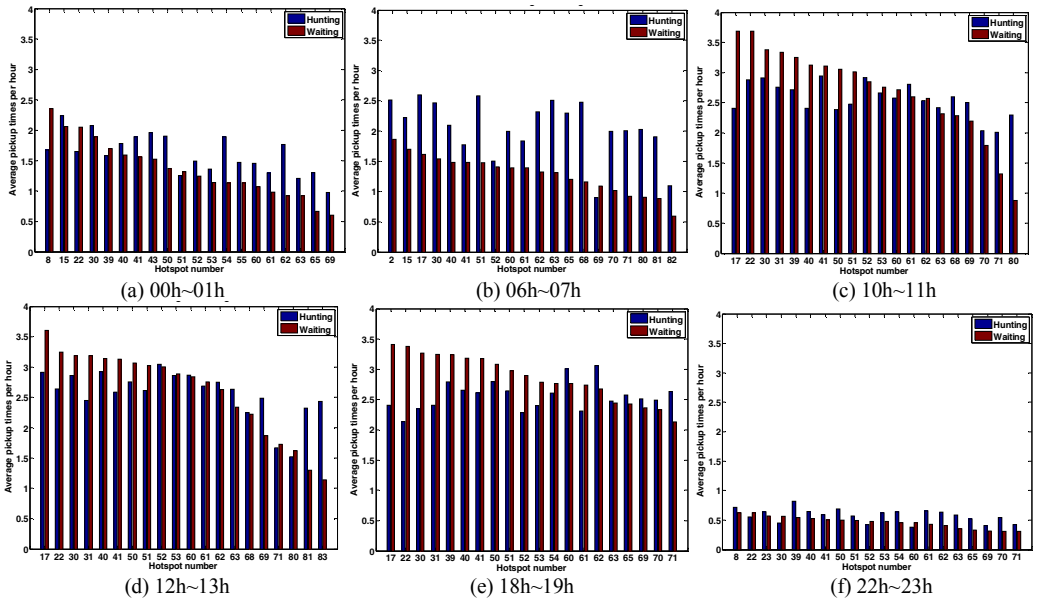

Figure 3. Comparision of the amount of passenger pickup beween the hunting and waiting behaviors

V. TAXI-PATTERN DISCOVERY

In this section, we are going to automatically extract taxi-patterns from the raw pickup/dropoff records to describe the semantics of taxi drivers' behaviors, and adopt a feature selection technique to select the underlying taxi-patterns that are most salient for taxi performance. There are three steps as follows: 1) we design a set of taxi-patterns in terms of times, locations, and the driving behavior before picking up or after dropping off a passenger, and obtain a rich collection

of features; 2) we use L1-Norm SVM to simultaneously select the most discriminative features and learn a predictor based on the selected patterns; 3) we predict the performance of a set of test taxis based on the selected patterns.

A. Passenger-Finding Strategy Descriptors

Normally, the factors that can significantly influence a taxi driver's performance are *subjective factors*. Subjective factors may determine the passenger-finding strategies when the taxi is unoccupied. Thus, in this paper, we focus on analyzing the passenger-finding strategies that the taxi driver adopts when the taxi is vacant. We find out two major passenger-finding strategies, which largely depend on the subjective factors, can significantly influence a driver's performance, they are

1) Hunting or Waiting before a Pickup? At a certain time and a certain location, which strategy should the taxi driver adopt, hunting passengers by keeping driving the taxi or waiting passengers by finding a hotspot nearby?

2) Local or Distance after a Dropoff: After dropping off a passenger at some time and some location, which strategy should the taxi driver adopt, finding passengers locally in the region where the last passenger gets off or returning to a region in a distance that the taxi driver is more familiar with?

To analyze these two subjective factors, first of all, we need to design a descriptor to describe the semantics of the two factors in machine-understanding formalism. To this end, we symbolize the both factors using a triplet

$$\text{(Time, Location, Strategy)} \tag{1}$$

For "Time", we divide one day into 12 equal intervals and the resulting time slots are {00h~02h, 02h~04h, …, 22h~24h}. For "Location", we use the region labels {1, …, 100} defined in Figure 1, in which each cell is approximately a 1200×1200 m^2 area. For "Strategy", it is a boolean value for indicating that a taxi driver is "hunting" or "waiting" passengers before a pickup or chooses "local" or "distance" to find the next passenger after a dropoff. This attribute can be described in terms of distance using the following rules:

$$\text{Pickup:} \quad d_{pick} \begin{cases} \le \tau_{pick} & \text{Waiting} \\ > \tau_{pick} & \text{Hunting} \end{cases} \tag{2}$$

$$\text{Dropoff:} \quad d_{drop} \begin{cases} \le \tau_{drop} & \text{Local} \\ > \tau_{drop} & \text{Distance} \end{cases} \tag{3}$$

where $d_{\text{pick}}/d_{\text{drop}}$ is the value in the last column of Table 1. The intuition is that: if the distance covered by a taxi before it picks up a passenger is not long, the taxi driver is very likely to wait passengers; if the distance covered by a taxi between the last dropoff and the next pickup is not long, the taxi driver is very likely to have found passengers locally in the region where the last passenger gets off.

By symbolizing the three attributes in (1) into taxi-pattern descriptors, we can obtain 12 (Time) × 100 (Location) × 2 (Strategy) × 2 (Pickup/Dropoff) = 4800 descriptors. The large collection of descriptors can enumerate all the possibilities of the passenger-find strategies. We construct a taxi-pattern table to count the times of a pattern a taxi driver has adopted during the selected 15 days. We have 4548 taxis and 4800 patterns, and then we have a 4548×4800 table. The element (i,j) denotes the times of jth patterns conducted by the ith taxi in the selected 15 days. The resultant table can thus be as a feature matrix for further data mining procedures.

B. Good/Bad Strategy Discovery

Next we aim to discover the underlying taxi-patterns that are most salient for working performance. We adopt a powerful supervised feature selection method, L1-Norm SVM [15], to simultaneously select the most discriminative features and learn a taxi performance predictor based on the selected patterns to discover which taxi-patterns are most associated with the drivers with better performance and which ones are most associated with worse performance.

The label of a taxi's performance is determined by the accumulated distance covered by the taxi in the selected 15 days when it is occupied. To assign a label to each taxi's feature (a row in the taxi-pattern table), we first sort the accumulated distance for all the 4548 taxis in descending order, and select the top 2000 taxis for our analysis. We use top 600 taxis out of the 2000 taxis as the positive examples and the bottom 600 taxis as the negative examples. We discard the rest 800 taxis with average performance. As a result, we have a data set with 1200 examples, in which 600 are positive and 600 are negative.

Based on the obtained training examples and their labels, we use L1-Norm SVM [15] to learn a classifier as well as select a small subset of most salient features for good and ordinary taxis from a collection of the taxi-patterns. L1-Norm SVM is a SVM with L1-norm regularization, which can lead to a sparsity solution over a large collection of features ("sparsity" means most features will get zero weights and they have no effect on the learned classifier).

The learning result of L1-Norm SVM is a set of feature weights $\{w_1,\ldots,w_M\}$, and the resulting classifier has the form

$$f(\mathbf{x}) = \mathbf{w}^{\mathrm{T}}\mathbf{x} + b = \sum_{i=1}^{M^+} w_i^+ x_i + \sum_{j=1}^{M^-} w_j^- x_j + b \tag{4}$$

where M^+ denotes the number of positive weights and M^- the number of negative weights. Due to the L1-norm regularization, the non-zero weights can be very few, i.e., $(M^+ + M^-) << M$, An intuitive interpretation based on (4) is that, for positive (or negative) weights, the corresponding features are more salient for the positive (or negative) examples, while for the zero weights, the corresponding features have no effect for discriminating positive and negative examples. Thus we can see L1-Norm SVM indeed performs a feature selection procedure implicitly during learning the classifier. The larger (or smaller) the feature's weight is, the more salient for the good (or ordinary) taxi.

We randomly split the 1200 examples into 800 training and 400 test examples for 10 times. The results we reported below are all in the form of "Mean±Std". The learning and test results are as follows: 186.4±7.6 taxi-patterns are selected (with non-zero weights) by L1-Norm SVM, in

which there are 135.3±6.8 positive features and 51.1±9.4 negative features. The test accuracy on the 400 examples is 0.853±0.013. We also compare the result obtained by a standard SVM based on the full set of features (i.e., 4800 taxi-patterns), and get an accuracy on the 400 test examples is 0.742±0.008. The prediction result based on the selected features clearly outperforms the one based on the all features, which implies that the selected features are indeed useful for suggesting efficient passenger-finding strategies.

The top 10 positive and top 10 negative taxi-patterns are listed in the descending order of the absolute value of the weights (see Figure 4). One can easily see that hunting is always a good way to identify the best taxis while waiting is always associated with ordinary ones. For [12h-13h, Region 69] and [0h-1h, Region 65], Pickup-Hunting are the good features, which is consistent with the fact that hunting is much better than waiting as shown in Figure 3. Although the pickup times is a bit smaller for hunting in [0h-1h, Region 39] an [12h-13h, Region 61] than waiting, in contrarily hunting is still a good behavior for both these regions. Possible explanation could be that during 0h~1h, hunting may happens not only in Region 39/61, but also in the downtown areas nearby, which together suggest hunting a good strategy.

	Top 10 Positive Taxi-Patterns
+1.083	[06h~07h, Region 100, Dropoff-Distance]
+0.783	[00h~01h, Region 61, Dropoff-Distance]
+0.745	[12h~13h, Region 69, Pickup-Hunting]
+0.695	[02h~03h, Region 53, Dropoff-Distance]
+0.643	[04h~05h, Region 100, Dropoff-Distance]
+0.629	[00h~01h, Region 51, Dropoff-Distance]
+0.571	[00h~01h, Region 39, Pickup-Hunting]
+0.545	[00h~01h, Region 65, Pickup-Hunting]
+0.529	[02h~03h, Region 67, Dropoff-Distance]
+0.512	[12h~13h, Region 61, Pickup-Hunting]

	Top 10 Negative Taxi-Patterns
-0.520	[18h~19h, Region 68, Dropoff-Local]
-0.513	[10h~11h, Region 14, Dropoff-Distance]
-0.463	[12h~13h, Region 30, Dropoff-Local]
-0.438	[10h~11h, Region 8, Dropoff-Local]
-0.422	[18h~19h, Region 58, Pickup-Waiting]
-0.401	[16h~17h, Region 46, Dropoff-Distance]
-0.393	[12h~13h, Region 16, Dropoff-Local]
-0.389	[12h~13h, Region 74, Pickup-Waiting]
-0.366	[08h~09h, Region 94, Pickup-Waiting]
-0.318	[18h~19h, Region 69, Dropoff-Local]

Figure 4. Top 10 positive/negative taxi-patterns discovered by feature selection. The numbers in the first column are the associated weights.

VI. CONCLUSION

In this paper we develop a novel method to represent the passenger-finding strategies using time-location-strategy triplet, transform the digital traces of 5350 taxis of Hangzhou in a year into a taxi-pattern table and training/test datasets, and select the determining taxi-patterns for efficient and inefficient passenger-finding via an advanced data mining algorithm. We find that the selected taxi-patterns can well interpret the empirical study results derived from raw data analysis and even reveal hidden "facts". Moreover, we have built a taxi performance predictor on the selected patterns and achieve a prediction accuracy of 85.3%. With the proposed method, for the first time we could mine most salient features for efficient and inefficient passenger-finding strategies, which can guide taxi drivers to perform better.

REFERENCES

[1] C.Ratti, R. M. Pulseli, S. Williams, and D. Frenchman "Mobile landscapes: Using location data from cell phones for urban analysis," Environment and Planning B: Planning and Design, Vol.33,No.5, pages 727-748, 2006.

[2] González, M. C., Hidalgo, C. A., and Barabási, A. L. "Understanding individual human mobility patterns," Nature,Vol. 453, pages 779-782, 2008.

[3] D. Ashbrook and T. Starner. "Using GPS to learn significant locations and predict movement across multiple users", Pervasive and Ubiquitous Computing, Vol.7, No.5, pages 275-286, 2003.

[4] K. Farrahi and D. Gatica-Perez "Learning and predicting multimodal daily life patterns from cell phones," In Proc. Of the 11th Int'l Conf. on Multimodal Interfaces and the 6th Workshop on Machine Learning for Multimodal Interface, pages 227-280, 2009.

[5] P. Mohan, V.N. Padmanabhan, and R. Ramjee "Nericell: Rich monitoring of road and traffic conditions using mobile smartphones," In Proc. Of the 6th ACM Conf. on Embedded Networked Sensor Systems, pages 323-336,2008.

[6] Wen Dong, and Alex Pentland A Network Analysis of Road Traffic with Vehicle Tracking Data, 2009, avaiable online: http://www.aaai.org /Papers/Symposia/Spring/2009/SS-09-04/SS09-04-002.pdf.

[7] Reades, J., Calabrese, F., Sevtsuk, A. and Ratti, C. "Cellular census: Explorations in urban data collection," IEEE Pervasive Computing, Vol.6, No.3, pages 30-38, 2007.

[8] Chang, H., Tai, Y., and Hsu, J.Y. "Context-aware taxi demand hotspots prediction,".Int. J. Bus. Intell. Data Min., Vol.5, No.1, pages 3-18,2010.

[9] Santi Phithakkitnukoon, Marco Veloso, Carlos Bento, Assaf Biderman, and Carlo Ratti. "Taxi-Aware Map: Identifying and predicting vacant taxis in the city," AmI2010, LNCS6439, pages 86–95, 2010.

[10] K.I. Wong, K.C. Wong, M.G.H. Bell, and H. Yang "Modeling the bilateral micro-searching behavior for urban taxi services using the absorbing markov chain approach," Journal of Advanced Transportation, Vol.39, pages 81-104, 2005.

[11] Hai Yang, C.S. Fung, K.I. Wong and S.C. Wong "Nonlinear pricing of taxi services. Transportation Research Part A: Policy and Practice," Vol.44,No.5, pages 337-348,2010.

[12] Liang Liu, Clio Andris and Carlo Ratti. "Uncovering cabdrivers' behavior patterns from their digital traces," Environment and Urban Systems, In Press, Corrected Proof, Available online 16 August 2010.

[13] Liang Liu, Clio Andris, Assaf Biderman, and Carlo Ratti "Uncovering Taxi Driver's Mobility Intelligence through His Trace," IEEE Pervasive Computing, 2009.

[14] Yamamoto, K., Uesugi, K., and Watanabe, T. "Adaptive routing of multiple taxis by mutual exchange of pathways", Int. J. Knowl. Eng. Soft Data Paradigm, Vol. 2,No. 1, pages 57-69, 2010.

[15] J. Bi, K.P. Bennett, M. Embrechts, C. Breneman, and M. Song, "Dimensionality Reduction via Sparse Support Vector Machines," J. Machine Learning Research, Vol. 3, pages 1229-1243,2003.