Uncovering cabdrivers' behavior patterns from their digital traces

Liang Liu *, Clio Andris, Carlo Ratti

Department of Urban Planning and Studies, Massachusetts Institute of Technology, MIT, 10-485, 77 Massachusetts Avenue, Cambridge, MA 02139, USA

A R T I C L E   I N F O

Article history:
Received 8 December 2009
Received in revised form 11 July 2010
Accepted 17 July 2010

Keywords:
Cabdriver
Operation behavior
Mobility pattern
GPS
Spatial–temporal data mining
Pervasive computing

A B S T R A C T

Recognizing high-level human behavior and decisions from their digital traces are critical issues in pervasive computing systems. In this paper, we develop a novel methodology to reveal cabdrivers' operation patterns by analyzing their continuous digital traces. For the first time, we systematically study large scale cabdrivers' behavior in a real and complex city context through their daily digital traces. We identify a set of valuable features, which are simple and effective to classify cabdrivers, delineate cabdrivers' operation patterns and compare the different cabdrivers' behavior. The methodology and steps could spatially and temporally quantify, visualize, and examine different cabdrivers' operation patterns. Drivers were categorized into top drivers and ordinary drivers by their daily income. We use the daily operations of 3000 cabdrivers in over 48 million of trips and 240 million kilometers to uncover: (1) spatial selection behavior, (2) context-aware spatio-temporal operation behavior, (3) route choice behavior, and (4) operation tactics. Though we focused on cabdriver operation patterns analysis from their digital traces, the methodology is a general empirical and analytical methodology for any GPS-like trace analysis. Our work demonstrates the great potential to utilize the massive pervasive data sets to understand human behavior and high-level intelligence.

1. Introduction

In recent years, the widespread deployment of pervasive computing technologies in cities has led to a massive increase in the volume of records of human spatio-temporal paths throughout the built environment. Collectively, these records allow for the development of new ways to study human behavior and provide new opportunities for the promising field of computational social science (Lazer et al., 2009).

Different types of data exist and can be used in the urban context. Some researchers have recently focused on the cellular network to study city dynamics and human mobility (González, Hidalgo, & Barabási, 2008; Ratti, Pulselli, Williams, & Frenchman, 2006; Reades, Calabrese, Sevtsuk, & Ratti, 2007). Others have been taking advantage of the data combinations produced by the overlay of digital information onto urban infrastructure – such as bus and subway routes, taxis, public utilities, and roads. Finally, much attention has been given to GPS-based data analysis, in order to understand individual outdoor movement, extract main destinations, predict movements and model routines (Ashbrook & Starner, 2003; Krumm & Horvitz, 2006; Liao, Fox, & Kautz, 2004; Liao, Patterson, Fox, & Kautz, 2005; Patterson, Liao, Fox, & Kautz, 2003). Many of these studies use aggregate data, and so far, few have used individual traces to better understand high-level human behavior and decisions, which would enable the creation of new services that autonomously respond to a person's unspoken needs.

This study aims to fill this gap in research through the analysis of an unprecedented database of individual, economically-driven spatio-temporal choices: a collection of records from 3000 cabdrivers, who take a total of 48 million trips covering 240 million kilometers for one year in Shenzhen, South China. We know the drivers are in competition to earn the most income via strategic operations patterns, but we are surprised to observe a noteworthy variation in ‘skill level’, as some drivers earn consistently up to six times more than others.

From this starting point, we refine our avenues of inquiry: What features of this very large database will lead to uncovering top drivers' strategy? How do these successful drivers optimize over the bounded resources of space and time? Or are these drivers profiting by deliberately choosing routes that are more costly for the customer, as every out-of-towner fears when stepping into a taxicab in an unfamiliar city? In solving these problems, we aim to develop a new methodology for the analysis of large amounts of spatio-temporal traces that, in the future, could be applied to other cities and domains.

Our contribution lies in the following:

(1) For the first time, we systematically study large scale cabdrivers' behavior in a real and complex city context (3000 taxis in a metropolitan area) through their daily digital traces. We identify a set of valuable features, which are
simple and effective to classify cabdrivers, delineate cabdrivers' operation patterns and compare the different cabdrivers' behavior.

(2) We develop a novel methodology and steps to spatially and temporally quantify, visualize, and examine cabdrivers' operation patterns. Drivers are categorized into top drivers and ordinary drivers by their daily income. We use the daily operations of 3000 cabdrivers in over 48 million of trips and 240 million kilometers to uncover the differences between top drivers and ordinary drivers: (1) spatial selection behavior, (2) context-aware spatio-temporal operation behavior, (3) route choice behavior, and (4) operation tactics.

The paper is structured as follows. First, we describe the raw data and our processing and feature extraction algorithm. Second, we propose a way to classify different cab drivers into high-earning 'top' drivers and average-earning 'middle' drivers (also referred to as average or ordinary) based on their income and rank. Third, we compare the different statistical operation patterns of top and middle drivers to reveal the 'invisible hand' of competitive driving strategy. Finally, we discuss our results and possible directions for future work.

2. Data sets

2.1. Data description

We have continuously collected data from 3000 taxi drivers' GPS traces for over 1 year (2008), yielding accurate data about taxi location in longitude, latitude form, timestamp, vehicle identification, operation status (empty or occupied), speed, and azimuth. The GPS traces include both occupied and empty running trips in order to describe the full operation behavior of cabdrivers. We also have contextual traffic condition information (car speed and volume), collected by the Shenzhen Transport Information Center and based on floating car and fixed point sensors.¹

Some information should be given about the modus operandi of taxicab drivers in Shenzhen. First, they do not work for a taxi dispatching system but simply for the hail market, meaning they cruise the road to find passengers. Most cars run 24 h a day in two shifts: the first shift (day) is from 7 am to 7 pm, the other (night) from 7 pm to 7 am. A single pair of drivers is linked to each vehicle, so we can actually consider each cab as a 2-person contained unit.

We should also mention that drivers do not have any significant spatial limitations; instead, they can operate freely in any part of the city and can choose their working zones. Finally, the fare structure for taxi drivers in Shenzhen is the following: 12.5 Yuan for the first 3 km and the 2.4 Yuan charge for each additional kilometer. This fare system varies in different parts of the world, and Shenzhen's system will certainly play a role in temporal operation decisions.

2.2. Initial data processing and feature extraction

After extracting important features from the GPS traces, we perform data cleaning, coordinate transformation, and map matching processes in order to understand cabdrivers' choices. In particular:

(1) Trip distance, travel time, travel speed, income/trip. For every trip, including occupied and empty legs, we calculated the trip distance on the road network, the corresponding travel time, travel speed and income. Of course, for trip fragments without a passenger, the income is zero.

(2) Origin/destination pair of each trip. For each trip, we estimate the origin (pick up location) and destination (drop off location) of the trip, hence gaining a good understanding of taxi travel demand at different times of the day. We consider origin to be a transition from empty to occupied in our timestamp sequence, and a destination to be a transition from occupied to empty in our sequence.

(3) Ratio of real path length over shortest path length (RRSL) and ratio of real path travel time over shortest path travel time (RRST). For each trip, given specific origin and destination, we calculate the shortest path length on the road networks and get the ratio of real trip distance over shortest path length (RRSL). Turner (2009) recently referred to this ratio as 'angularity' when analyzing courier GPS traces. At the same time, we also estimate the travel time, given the real time traffic speed on the road links if the driver uses the shortest paths, then we calculate the ratio of real path travel time over shortest path travel time (RRST).

Jiang, Yin, and Zhao (2009) look at taxi cab travel on a large road network over 6 months. In comparison, they simulate a Levy Flight process, where the of steps in a walk are chosen randomly from a fat tailed distribution of directions (e.g. left, right, up, and down) and mimic the degrees of freedom provided by the road length infrastructure. Their findings show that real data from 50 taxis exhibit this flight behavior because of the constraints of the underlying street network.

3. Classification of taxi drivers

3.1. Calculation of taxi income

Our first analysis is based on the calculation of the daily income of taxi drivers. We derive this measure from the fare structure discussed in Section 2.1, and our calculations of trip length on the road grid, and vehicle status (occupied/empty). We find that the average daily income follows a normal distribution for all taxi drivers for 1 year (Fig. 1).

Most drivers (ordinary drivers) garner income are around the mean value (700–800 Yuan, approximately 110 US$), several are very high (top drivers, 1250 Yuan), several very low (200–300

Fig. 1. The daily income of different taxi drivers shows a normal distribution (Unit: Yuan).

Yuan). What is very surprising is the variance of this distribution: as mentioned, top drivers earn up to six times more than low performing drivers!

How could this be? Perhaps top drivers, despite having the same shift length (7 am–7 pm or vice versa) work more intensely and take shorter pauses during their shifts? This is partially true: top drivers work for longer hours (20 h), than ordinary drivers (18 h) and bottom drivers usually work 12 h and take one shift off. However, even when data are normalized based on working times, the top drivers’ performance is still remarkable in term of hourly wage, 150% of the average.

3.2. Is top cabdrivers’ performance consistent?

We now ask if the top cabdrivers’ excellent performance is consistent (indicating precision), or if these driver rise to the top of the income chart simply by chance of a lucky workday.

In order to answer this question, we calculate the average rank and income for each cab driver and rank these results. Through this process, we eliminate the random error and select the real top drivers to analyze their operation patterns. We estimate both the average daily income and the overall rank for all the 3000 drivers to better define the two classes of drivers:

(1) Top drivers: average daily income greater than 900 Yuan, and daily rank very consistent (standard deviation <30);
(2) Ordinary drivers: average daily income less than 900 Yuan, and daily rank very inconsistent (standard deviation >100);

We call the driver with the overall average rank of 1—first place—the No. 1 Driver. This driver keeps an incredible record: over one year, he ranks as the number one driver 44% of the time. Seventy-six percent of the time he is in the top three taxi drivers, and 88% of the time he is in the top 10. The phenomenon of the top driver can be seen with a number of top drivers, so consequently, we see that the operational skill is not random, since its top drivers make better decisions, and that those who master the art of driving customers around the city have a much more profitable workday.

Now that we have “singled out” one cab out of many, we think of privacy. Indeed, with the advent and use of tracking devices comes a new set of privacy ethics. We have spotted the “number 1” driver and his location at different timestamps. The choices of the top driver, and the other drivers are exposed in this way. Drivers are marked with an ID, but we changed this ID for each driver so the driver’s name could not be identified. We think these types of studies are interesting starting blocks for discussing the ethics of privacy regulations, as marked by what kind of personal information is exposed.

So what is the operation pattern of the top drivers? And what can these secrets tell us about their high-level behavior and decisions?

4. Operation patterns comparison: top drivers vs. ordinary drivers

In this section we compare several operation patterns of top drivers and ordinary drivers to answer why top drivers perform much better than ordinary drivers: (1) spatial selection behavior, (2) context-aware spatio-temporal operation behavior, (3) route choice behavior, and (4) operation tactics.

4.1. Spatial selection patterns

Location is one of the key issues for many businesses. Is location important for cabdriver operation? Yes, our detailed analysis conveys strongly that location does matter.

We discover the contour lines that divide cabdriver pickup point distribution, which elaborately delineates the spatial distribution of two key operation areas of taxicabs: 1—Luohu and Futian District; 2—Nanshan District (Fig. 2a). Note that the average speed is 36 km/h in Nanshan District, 18 km/h in Futian District and 15 km/h in Luohu District.

In order to uncover the difference between top drivers and ordinary drivers, we analyze their spatial selection patterns. Through counting the proportion of different operation districts for the top and ordinary taxi drivers, we are able to understand how top and ordinary drivers select their areas of operation. This intuitive comparison highlights the different spatial selection patterns of top drivers and ordinary drivers (Fig. 2b). Here, these two different areas are not distinct to only top drivers or only ordinary drivers, but represent a significant clustering of an unusual preponderance of the two respected types.

At the same time, a detailed comparison of average speed rate, the idle rate, average income rate and ratio of top drivers of each district is shown in Fig. 2c. In this figure, Average Speed Ratio, Idle Rate ratio and Average Income Rate are defined as follows:

(1) Average speed ratio. In the city of Shenzhen, the design speed of roads is 60 km/h, while in reality, most roads cannot achieve the maximum speed because of different road conditions such as the traffic signal and traffic congestion. So we compute the average speed of each district and divide it by maximum speed. The value of average speed ratio is between 0 and 1.
(2) Idle rate. The proportion of idle taxis to total taxis in that area. The idle rate describes the competence of the taxis in a specific area. The value is between 0 and 1. This feature describes the competitiveness of taxi operation.
(3) Average income rate. The ratio of ordinary drivers’ daily district average income to top drivers’ daily district average income. The average income rate describes by how much top drivers outperform the ordinary drivers in different districts.

Fig. 3c shows that the Nanshan district is a focal operation area for top drivers, and their choice proves to be more lucrative than the choices of ordinary drivers who focus on the Luohu and Futian areas. If we combine this with the average speed, idle rate and average income rate in different districts, it is easy to understand why top drivers prefer to take roads with higher speeds, less competitive and more profitable locations. The statistical results show that top drivers focus on their earning ability per time unit.

4.2. Optimal operation space time path

Is there an optimal operation space time path for different cabdrivers? In order to answer this question, we calculate the correlation coefficient of top drivers and ordinary drivers. The correlation coefficient $\rho_{XY}$ between two random variables $X$ and $Y$ with expected values $\mu_X$ and $\mu_Y$ and standard deviations $\sigma_X$ and $\sigma_Y$ is defined as:

$$
\rho_{XY} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}
$$

where $E$ is the expected value operator and $\text{cov}$ is covariance.

Each individual taxi driver user $i$ has the trace sequences $U_i(x_i, y_i, t_i)$, where $x_i, y_i, t_i$ represents the x coordinate, y coordinate and time coordinate of user $i$ in the $j$th space time point.

The sets of user trace sequence compose the space time path (Hagerstrand, 1970). Let $X(x_i, y_i, t_i)$ represent the sequence of space time points of user $X$, and $Y(x_i, y_i, t_i)$ represent the sequence of
space time points of user Y. From these, we can obtain the correlation of top drivers’ space time path by using Eq. (1).

If we correlate the top taxi drivers with their top 200 TAZs (Traffic Analysis Zone), we find the following rule: the greater the correlation coefficient, the closer rank to the No. 1 position (62% top drivers’ correlation coefficients are surprisingly greater than 0.75). That means the more a driver can muscle on the top ranked driver’s spatial–temporal distribution of operation zones, the more
revenue he can make. On the contrary, the correlation coefficient of ordinary drivers is quite small (less than 0.60), which means they are not similar to one another.

This is very interesting, as it seems there is an optimal operation space time path that could be the operation guidance for the taxi drivers. So let us look again in more detail at the top cabdrivers’ operations.

4.3. Comparison of spatial–temporal operation

Cabbdrivers operate in both spatial and temporal scales, meaning that location is not the only consideration, but that time matters as well. In order to investigate different operation patterns of top and ordinary drivers in a more detailed way, we compare the spatial–temporal distribution of top drivers and ordinary drivers, we...
calculate the operation centroids over different time periods, and compare the top drivers with the ordinary drivers. From these, we can get a sense of what determines the different drivers’ focal areas.

4.3.1. Cluster algorithm

In order to obtain the different taxi drivers’ operation characteristics, we use a K-means clustering algorithm to calculate operation centroids (cluster center of operation points):

First we retrieve all the operation points \( P_i(x_i, y_i, t_i) \) in a day, where \( x_i \) is the horizontal axis, \( y_i \) is the vertical axis, and \( t_i \) is time axis (unit: second).

Second we divide the cluster number \( j \) into daily segments, where one day is divided into four categories: early morning, morning, afternoon and evening, namely, \( j = 1, 2, 3, \) and 4.

Third we compute a cluster center \( P_j \) for each cluster, ensuring that the sum of \( |P_i - P_j|^2 \) is minimized, and the distance between two points is the Euclidian distance.

4.3.2. Cluster results

According to the K-means algorithm, we can obtain the operation centroids for different time periods for the top drivers and ordinary drivers (in Fig. 3a and b, the four numbers represent \( x \) coordinate, \( y \) coordinate, time and TAZ).

From Fig. 3a, we learn that there is an obvious spatial–temporal routine in the way that top drivers operate. Before 6 am, they generally operate in the Guomao and Dongmen (106) districts. From 8 am to 9 pm, they operate on the center of Nanshan district and Science Park (328, 335), and from 9 pm to 12 pm, they operate around Huqiangbei District (208).

If we consider the transportation congestion conditions in different districts (Luo Huo, Futian and Nanshan), it seems that top drivers have an excellent trade-off between travel demand and road conditions in determining their operation zones. During the day, when traffic is congested, they choose to operate in the Nanshan district where the traffic is lighter. Even though the travel demand in Nanshan is less than that of the Luohu-Futian district, they can satisfy their operation requirement nevertheless. Meanwhile, during the evening and early morning, top drivers operate in the Luohu-Futian district where the travel demand is greater, because during this period, the traffic is lighter in this area. Uncovering these econometric conditions gives us an added dimension to understanding the strategies and motivations of top performers. Similarly, we can now reveal some of the shortcomings of drivers who are slightly less successful.

Indeed, the skill of ordinary drivers is less competitive. The operational scale of ordinary taxi drivers is much smaller than that of the top drivers (Fig. 3b). They, perhaps naïvely, operate in the Central Business District (Zone 103: Guomao; Zones 105 and 106: Dongmen) during the early morning, morning and evening, while in the afternoon, they operate in Huqiangbei and Shangbu areas (217). An ordinary driver focuses more on the traditional city center—Luohu-Futian Area, preferring to stay in the area with the most travel demand, but ignores the potentially-detrimental impact of congested traffic conditions. Working backwards, this elastic balance between passenger supply and traffic conditions is best explained by the behavior of the top drivers (Fig. 3c–f).

4.4. Operation tactics

4.4.1. Trip distance, travel time, speed

From a cabdriver’s activity history, we can reveal the different operation patterns of two different classes of drivers. When carrying a passenger, the trip distance, trip time and income of top drivers are significantly higher than ordinary drivers. While idle, the trip distance and trip time of top drivers are significantly lower than that of ordinary drivers, e.g. the top drivers do not search around as much when without a fare. But whether in operation or idle, the speed of top drivers is much higher than that of ordinary drivers. It seems that top driver are very good at both saving a lot of time and generating more business. Table 1 shows the results.

In Table 1, operation distance means the travel distance between the passenger pick up point and drop off point; while idle distance means the travel distance between the last passenger drop off point and the next pick up point. The data in Table 1 is the average value in the 1 year observation period.

### 4.4.2. Ratio of real path length over shortest path length (RRLS) and ratio of real path travel time over shortest path travel time (RRST)

For each trip, we calculate the real path length and shortest path length, real path travel time over shortest path travel time, and find interesting patterns showing that top drivers and ordinary drivers have different route choice patterns according to the fare structure (Fig. 4).

From Fig. 4, we can see the different operation patterns of top drivers and ordinary drivers. Under 3 km, top drivers’ RRSL is very close to 1, indicating that their route choice is very close to the shortest path, however, after 3 km, the RRSL increases, which means they may take more indirect routes. In contrast, the ordinary drivers do not take the shortest path under 3 km, and above 3 km, their indirectness is less than that of top drivers.

If we consider the travel time, top drivers’ real travel time is less than the time it would take if she were to take the shortest path, and RRST decreases when distance increases, which means top drivers are skilled in finding the fastest way rather than shortest path because, perhaps, they are more familiar with the real time traffic conditions of the city. In contrast, ordinary drivers usually take more time in traveling between points than the shortest path would ordinarily require.

Combining these two factors, it seems that top drivers earn more money not because they take passenger on longer routes but because they work faster—taking passengers through the city by the fastest, least congested routes even if this causes a large amount of detours. It perhaps demonstrates that top drivers have more knowledge about the city, such as traffic conditions and street network. They do not navigate according to the least change (Hiller & lid, 2005). At the same time, they do not care much about length, which is shown by the steep change at 3 km.

In order to compare the differences of top drivers and ordinary drivers in term of RRSL and RRST feature, we calculate the statistics of these two features in three different districts (in Fig. 5).

Fig. 5 shows top drivers’ RRSL increases from Luohu, Futian to Nanshan, while their RRST decreases in these three districts, which means that top drivers use less time to drive more (make more money) in Nanshan district. On the contrast, ordinary drivers’ RRSL

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Status = operation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item</td>
<td>Status = operation</td>
</tr>
<tr>
<td>Top drivers</td>
<td>Distance (km)</td>
</tr>
<tr>
<td>6.50</td>
<td>11.66</td>
</tr>
<tr>
<td>Ordinary drivers</td>
<td>4.86</td>
</tr>
<tr>
<td>Difference</td>
<td>1.64</td>
</tr>
<tr>
<td>Percent change</td>
<td>33.63</td>
</tr>
<tr>
<td>Status = idle</td>
<td></td>
</tr>
<tr>
<td>Top drivers</td>
<td>Distance (km)</td>
</tr>
<tr>
<td>3.12</td>
<td>10.14</td>
</tr>
<tr>
<td>Ordinary drivers</td>
<td>5.22</td>
</tr>
<tr>
<td>Difference</td>
<td>–2.1</td>
</tr>
<tr>
<td>Percent change</td>
<td>–40.23</td>
</tr>
</tbody>
</table>

4.4.3. Ratio of real path travel time over shortest path travel time (RRST)
and RRST increase from Luohu, Futian to Nanshan. That means they take more time to drive the same distance (make the same amount of money) as top drivers. Their earning ability per unit time is less than that of top drivers.

In order to compare the different driving behavior of top drivers and ordinary drivers at different distances, we draw the distance–time relationship diagram (Fig. 6) and see that, usually, top drivers drive faster than ordinary drivers below 3 km, and drive even faster than ordinary drivers beyond 3 km. Because top drivers seem to be more familiar with the street network and traffic condition, they may have more flexibility during the longer trip.

Statistics show that for trips around 3 km, top drivers spend on average 1 min less than ordinary drivers, while for trips beyond 3 km, top drivers spend 2 min less than ordinary drivers. At the same time, top drivers average 15 min less than ordinary drivers when idle. In sum, top drivers spend 16–17 min less than ordinary drivers for every single operation cycle (operation trip plus idle trip), which allows top drivers to pick up around 25–30 more fares than ordinary drivers.

5. Conclusions and outlook

In this paper we develop a novel methodology to understand cabdrivers’ operation patterns by analyzing their continuous digital traces. For the first time, we systematically study large scale cabdrivers’ behavior in a real and complex urban context (3000 taxis in a metropolitan area). We identify a set of valuable features, such as RRSL and RRST, which we found to be straightforward and effective for cabdriver classification, cabdrivers’ operation patterns delineation and different cabdrivers’ behavior comparison. We thought that top cabdrivers would try to optimize length of trip; instead they put more emphasis on the time of trip, and they want to maximize their usage of time and finish their tasks as soon as possible. Thus top drivers usually have higher speed than ordinary drivers both on operation status and idle status. Because top drivers have better knowledge of street networks and traffic condition, they have more flexibility during the longer trip.

We look forward to real time optimization and feedback: could the overall spatial–temporal demand data be shared with all taxis so that they can better cover demand areas? Could everyone benefit from the intelligence of the top driver, following him in order to increase their revenue? Could we design a better system for providing real time feedback to individuals to optimize system performance? This is where pervasive computing can allow us not only to better understand our cities and inform policy-makers, but also to improve them with the help of urban planners.

What does this analysis have to do with the city? What does it mean for the city? This analysis is interesting from a mode choice and taxi demand point of view. It is also interesting for city planners because it may dispel notions about where taxi drivers find the highest need. An expected guess would be that taxis find their

---

**Fig. 4.** Top and ordinary drivers’ RRSL (left) and RRST (right) for various distances.

**Fig. 5.** RRSL and RRST comparison for top drivers and ordinary drivers in different districts.

**Fig. 6.** Comparison of driving behavior for top and ordinary drivers along the distance.
most profit in the Central Business District, but the observations showed that high-earning taxis made their profits in another part of the city. This may come as a surprise to planners, who may react by building taxi stands and safe standing areas, re-routing busses, or posting taxi company phone numbers for potential patrons. Of course this fragile economic balance would need further consideration.

It is also interesting to infer activities from the temporal resolution of the traces. When do conference-goers need a ride from the airport to a hotel? Do they arrive on Monday mornings and stay for three days? What does this mean for convention center placement? Which attractions are being visited by tourists via cab? Do people seem to be choosing taxi service for their daily ride to work? Are nightlife institutions producing a significant number of nighttime calls for rides home? These questions are helpful for a city because they describe resource and space usage.

These findings are inspiring because we have various assumptions about human behavior in the social life but seldom get the direct data sets to prove or defend them. Studying cabdrivers’ operation behavior from many GPS traces opens the new venue for research about human behavior. We believe this type of analysis and methodology is replicable and well-suited for a number of application domains like logistics, mass transit, pedestrian initiatives, or zoning. We are excited to see that pervasive data sets can help us understand human behavior and motivational intelligence.

References


