Turning meter transactions data into occupancy and payment behavioral information for on-street parking

Shuguan Yang^a, Zhen (Sean) Qian^{a,b,*}

^a*Department of Civil and Environmental Engineering
Carnegie Mellon University, Pittsburgh, PA 15213, U.S.A.*

^b*Heinz College
Carnegie Mellon University, Pittsburgh, PA 15213, U.S.A.*

Abstract

Over 95% of on-street paid parking stalls are managed by parking meters or kiosks. By analyzing meter transactions data, this paper provides a methodology to estimate on-street time-varying parking occupancy and understand payment behavior in an effective and inexpensive way. We propose a probabilistic payment model to simulate individual payment and parking behavior for each parker. Aggregating the payment/parking of all transactions leads to time-varying occupancy estimation. Two data sets are used to evaluate the methodology, parking spaces near Carnegie Mellon University (CMU) campus, and near the Civic Center in San Francisco. The proposed model generally provides reliable estimations of occupancies at a low error rate and substantially outperforms other naive models in the literature. From the results of the experiments we find that people generally tend to slightly underpay in CMU area, whereas for Civic Center area, payment behavior varies by time of day and day of week. For Fridays, people generally tend to overpay and stay longer in the mornings, compared to underpaying and parking for shorter durations in the late afternoons. Parkers' payment behavior, in general, is more variable and noisier around Civic Center than around CMU. Moreover, we explore the effective granularity, defined as the highest spatial resolution for this model to perform reliably. For CMU areas, the effective granularity is around 10-20 spaces for each block of streets, while it is 150-200 spaces for the Civic Center area due to more random parking behavior.

Keywords: Time-dependent parking occupancy, payment behaviors, parking meter transactions

1. Introduction

With the boom of global vehicles ownership (Sousanis 2011), the last few decades witnessed the emerging of parking supply scarcity (Outlook 2011). For most areas with dense population, the amount of parking spaces, especially on-street parking stalls, are insufficient. A significant portion of the total traffic is cruising for parking during business hours in some

*Corresponding author

Email addresses: shuguany@cmu.edu (Shuguan Yang), seanqian@cmu.edu (Zhen (Sean) Qian)

1
2
3
4
5
6 urban areas (Shoup 1997). To understand parkers' behavior and optimize the utilization of park-
7 ing resources, one needs to know precisely parking occupancies in high spatial and temporal
8 resolution, and how much travelers pay for on-street parking. Understanding payment behav-
9 ior is critical for both congestion mitigation and parking enforcement. Currently, most of the
10 on-street parking occupancy data is collected by networks of parking sensors (SFMTA 2014,
11 Hsu 1989), which are costly to install and maintain, not practical for scaling up to large-scale
12 networks. Even with those sensors, we do not have a good understanding of travelers' payment
13 behavior, how much (or whether) they pay versus how long they actually park.
14

15
16 One ubiquitous data source that can be used to infer payment behavior and parking oc-
17 cupancy is parking meter/kiosk transactions. Over 95% of on-street paid parking stalls are
18 managed by parking meters or kiosks. However, meter data has not been well exploited in the
19 literature (Ottosson et al. 2013). By analyzing meter transactions data, this paper provides a
20 methodology to provide estimates for on-street parking occupancy and understand payment be-
21 havior in an effective and inexpensive way. We propose a probabilistic model to simulate park-
22 ing payment behavior for individual parking transactions. Aggregating the payment/parking of
23 all transactions leads to real-time occupancy estimation. We also examine the spatial granular-
24 ity for such estimations by considering the tradeoff between estimation reliability and spatial
25 resolutions.
26
27

28
29 Occupancy monitoring is critical for parking management. The initial motivation of *SFPark*
30 project is maintaining on-street parking occupancy rates within an "optimal" range (60% to
31 80%) by adjusting parking rates accordingly (SFMTA 2014). Moreover, time-dependent and
32 fine-grained occupancy information is required for various network models regarding parking
33 choice (Chen et al. 2011), parking rate (Qian & Rajagopal 2014, 2015, Lei & Ouyang 2017,
34 Pu et al. 2017), traffic congestion (Arnott & Inci 2006), parking policies (Cullinane & Polak
35 1992, Arnott 2006, Verhoef et al. 1995). Also, reliable estimations of real-time parking occu-
36 pancies are serving as an essential input module for numerous transportation systems, such as
37 park and ride system (Rodier & Shaheen 2010, Chen et al. 2014) and free-floating car sharing
38 mode (Balac et al. 2017). Among various types of urban parking, on-street parking is a conven-
39 ient yet limited option. Unlike garage parking, where real-time occupancy data can be obtained
40 via monitoring the total number of vehicles entering and exiting the garage, on-street parking
41 needs physical spot-by-spot sensors to obtain accurate occupancies, such as infrared distance
42 meter and magnetometer-based sensor (Hsu 1989). As part of the *SFPark* project, 8200 in-
43 ground wireless magnetometer-based sensors were installed in pilot areas (SFMTA 2014) for
44 each managed parking spot. Although these sensors have gathered valuable occupancy ground
45 truth, the excessive cost of installing and maintaining the sensor networks has limited its ap-
46 plication to large-scale networks. Apart from in-ground sensors, there are sensing approaches
47 based on smartphones (Nawaz et al. 2013) and on-vehicle sensors (Mathur et al. 2010), all of
48 which impose additional cost and dependencies onto the current parking systems.
49

50
51 In practice, on-street parking meters/kiosks have been widely used to manage parking spaces.
52 The parking transactions taking place in those meters/kiosks contain information regarding
53 when, where and how long a parker pays for parking. This information can be used to infer
54 parking occupancy without deploying additional sensing systems. However, this idea does not
55 seem to be fully explored in the literature. Most studies work with the aggregated total pur-
56 chased parking time, omitting the individual information provided by each transaction. As part
57
58
59
60
61
62
63
64
65

of the evaluation of *SFPark*, a sensor independent rate adjustments (SIRA) method is proposed, in which a linear regression model based on cumulation of purchased parking time is used to estimate on-street parking occupancy (Pierce & Shoup 2013, Millard-Ball et al. 2014). This model is applied to occupancy estimation on a monthly basis. It cannot estimate time-varying occupancy, and the estimation error increases significantly as the time resolution goes up. Similar issues also exist in a case study in Seattle (Ottosson et al. 2013), in which occupancies are calculated by simply cumulating all transactions and are used to evaluate the impact of rate changes.

To obtain a reliable estimation of occupancy from meter transactions data, we need an understanding of individual parking payment behavior. Simply assuming a parker leaves exactly at the time of ticket expiration may lead to substantial estimation errors. The problem of estimating occupancy in high spatial and temporal resolutions is non-trivial due to unpaid, overpaid and underpaid parking. In the study of price-sensitive parking demand, Ottosson et al. (2013) discovered that parking occupancies estimated directly from aggregating transactions (namely every parker leaves exactly at the time ticket expiration) are generally lower than their actual observations, and such discrepancies are substantial in some cases. A similar conclusion was reached in a study comparing data from parking stations and counted data for major paid parking blocks in Seattle (Seattle 2011*b,a*), where the estimation error varies from 6% to 55%, with occupancies derived from transactions always being lower than the counted data. The issue indicates that many travelers often underpay and stay longer than the period they bought for, or simply do not pay. This exists in other cities as well (Weinberger et al. 2010). Another survey conducted in Seattle (SEATTLE 2002) shows that on average drivers parked 6 minutes to two hours longer in several neighborhoods provided with a 2-hour parking limit. Millard-Ball et al. (2014), Pierce & Shoup (2013) also discover that drivers' parking payment behavior is influenced by a set of external factors including time of day, day of week and length of parking period altogether. There are also factors that can dominate the occupancy estimation in certain cases, such as unpaid parking, overnight parking and trucks' roadside standing. It is critical to model individual parking payment behavior and incorporate external factors for reliable occupancy estimation.

This paper proposes a probabilistic model for each individual parker/transaction, and estimates the time-varying occupancy at the block level. We also conduct various experiments to verify the performance of our model and explore payment behavior varying by time of day, location and granularity. Effective granularities for reliable estimations are discussed. In this study, the granularity is defined as the spatial resolutions for occupancy estimation. A higher granularity implies the occupancy estimates over a smaller group of parking spaces than a lower granularity. Naturally, both parkers and parking managers would prefer the highest possible spatial resolution for parking occupancy estimation. However, parking occupancy can be sometimes dominated by the nature of random parking payment/choices within blocks. At a very high granularity/resolution, the occupancy estimate can be very challenging. Therefore, it is necessary to examine the proper granularities for the estimation model to work best for balancing estimation accuracy and estimation resolution.

The remainder of this paper is organized as follows. Section 2 discusses the individual payment behavior model and the method to estimate real-time occupancies. The dataset used in this study is briefly described in Section 3. Section 4 discusses several experiments regarding

the performance of the model under various conditions and the effective granularity for the model to be reliable. Section 5 concludes this paper.

2. The Model

In this section, an individual payment behavior model is proposed. The methodology for learning the behavior model and estimating time-dependent on-street parking occupancy, along with an adjustment for overnight parking, is discussed. Then, the corresponding solution algorithm is proposed.

2.1. Modeling departure time from parking and assumptions

2.1.1. The general idea

Individual parkers may not conform to the schedule of the parking sessions they purchase in advance. Some would leave before their parking sessions expire, in other words, they overpay for their parking, while some others leave later than the session expiration time, namely they underpay. There are people who park without paying, which also contributes to occupancy. Moreover, the parking and payment behavior differ from driver to driver, varying substantially by time of day, day of week, meter rates and locations. For example, travelers may be less worried about parking violation when parking in those areas with loose enforcement. As a result, they are likely to underpay in those areas.

The idea of this whole methodology is that the probability for an individual parker underpay, overpay and no-pay for parking can be modeled. The actual departure time from parking of an individual parker follows a probability distribution. The distribution is dependent on the ticket expiration time, purchased duration, time of day, day of week, and locations. Those probability distributions, when aggregated over all transactions/parkers, can reproduce the actual occupancies observed in the real world. We will use actual occupancy data observed on sampled days to calibrate the probabilistic model, which are further used to estimate and predict occupancy.

2.1.2. Choice of Distributions

In literature, although various distributions have been applied to represent the distribution of parking duration, such as negative exponential distribution (Richardson 1974, Cleveland 1963) and Gamma distribution (Cao & Menendez 2013), little research have been conducted regarding the probability distribution of departure time given the parking payment information for on-street parking.

In this study, we model the departure time of an individual vehicle by a logistic distribution. The reasons for using a logistic distribution are as follows. First, it exhibits a relatively low probability density of the vehicle leaving at the beginning of a parking session, and a higher probability density as time progresses until the expiration time, which is consistent with the parking payment practice. Second, the scale s of the distribution can be used to control the level of dispersion, as well as the tail shape of the distribution. We can utilize data to learn the scale s for a good fit of the distribution. Third, it has a simple form of probability density distribution (PDF) and cumulative density distribution (CDF), which brings computational advantages.

2.1.3. Assumptions

A set of assumptions regarding drivers' payment behavior are made throughout this paper, and they are summarized as follows,

1. As discussed in Section 2.1.2, the parking departure time is assumed to follow a logistic probability distribution. For all parkers who arrive at the same time of day and day of week, at the same location, and with the same purchased parking duration, the distribution of departure time is identical.
2. The mean and scale (variance) of the probabilistic distribution of the departure time is linear with respect to the purchased duration and expiration time of the purchased parking session.
3. The payment behavior of a parker is independent of other parkers and prevailing occupancies.
4. The exogenous factors influencing parking payment behavior include time of day, day of week and parking location. Generally, parkers have different payment and parking behavior when using spaces at different time of day and locations.
5. All overnight parked vehicles leave before noon. This assumption is made after analyzing parking ground truth as well as payment transactions on the two sample data sets used in this study. The discrepancy between the estimated occupancy and ground truth seems constant from the early morning to noon.

2.1.4. Model formulations

For an individual parker, suppose that he/she purchases a ticket of parking duration l , and expires at time μ , then the starting time of the parking session can be expressed by $\mu - l$. This parker overpays if he/she leaves prior to μ and underpays if after μ , as the payment behavior of the parker demonstrated in Figure 1. The probability that this parker is still parked at time t , conditional on purchasing a ticket that expires at μ of duration l , is denoted by $P(t)$, and can be expressed as:

$$P(t|\mu, l) = 1 - F(t|\mu, l) = 1 - \int_{-\infty}^t f(\tau|\mu, l) d\tau \approx 1 - \int_{\mu-l}^t f(\tau) d\tau \quad (1)$$

where $F(t|\mu, l)$ is the cumulative distribution function of the actual departure time conditional on a transaction μ, l , and the corresponding probability distribution function is denoted by $f(t|\mu, l)$, representing the probability of the parker departures at time t . In practice, parkers could not leave before the starting time of their parking session $\mu - l$. As a result, we adopt a probability density distribution such that the probability of the departure time being earlier than $\mu - l$ is very close to zero and thus negligible.

Place Fig 1 here.

By Assumption 1, we assume that $f(\cdot|\mu, l)$ follows a parameterized logistic distribution with mean μ_f , determined by (μ, l) , and scale s , determined by l ,

$$f(t|\mu, l) = \frac{e^{-\frac{t-\mu_f(\mu, l)}{s(l)}}}{s(l) \cdot (1 + e^{-\frac{t-\mu_f(\mu, l)}{s(l)}})^2} \quad (2)$$

Assumption 3 ensures that the probability distribution function is dependent on the time of day t , but is independent on the prevailing parking occupancy. By Assumption 4, $f(t|\mu, l)$ will be tuned for any time of day t , a specific weekday and a specific location.

For a single parking transaction, the mean of the departure time distribution μ_f can be seen as an indicator of the expected payment behavior (see Figure 1 for an example, this parker is more likely to underpay than overpay), while the scale s describes the variance of the actual departure time. By intuition, the longer a vehicle parks, the later it leaves, and the more uncertain (namely more variance) its departure time is. By Assumption 2, μ_f and s of the distribution change linearly with respect to l , the duration of the parking session, shown as the following,

$$\begin{cases} \mu_f(\mu, l) - \mu = a \cdot l + b \\ s(l) = c \cdot l + d \end{cases} \quad (3)$$

where $(a, b, c, d) = \omega$ is the parameter set of the individual payment behavior model. Each of the four parameters depicts a particular aspect of the vehicle's parking behavior:

- a reveals the influence of the purchased parking duration on the actual departure time.
- b stands for a shift of the actual departure time from the parking session expiration time.
- The symbol of $(a \cdot l + b)$ indicates the overall expectation of an individual parking behavior being either overpaying or underpaying. A negative number leads to an overall overpaying behavior if aggregating over all such travelers, where a positive number represents the tendency of underpaying.
- c and d are controlling the variance/uncertainty of the actual departure time. c is measuring the sensitivity of departure time variance to purchased parking duration, and d is the offset of the variance. The larger c and d are, the more uncertain an individual parker's departure time would be.

Thus, $P(t|\mu, l)$, the probability of the parker being still parked by time t , can be written as,

$$P(t|\mu, l) = 1 - F(t|\mu, l) = \frac{1}{1 + e^{\frac{t - \mu_f}{s}}} = \frac{1}{1 + e^{\frac{t - \mu - a \cdot l - b}{c \cdot l + d}}} \quad (4)$$

The hazard rate $h(t)$, the probability of a parker leaving between t and $t + dt$, conditional on μ, l and not having left at time t , can be calculated as $h(t) = f(t|\mu, l)/P(t|\mu, l)$.

2.2. Occupancy estimation

Based on the proposed payment behavior model for individual parkers, we derived a method for estimation of parking occupancies. In short, parking occupancy for a given time point can be estimated by aggregating the probabilities of each vehicle still parking by then. Moreover, by analyzing real-world data sets, we find it important to incorporate overnight parking when estimating occupancies for morning hours. Thus, an adjustment for estimation is made for overnight parking in Section 2.3.

In this study, we interpret parking occupancy as the total number of parked vehicles at a particular time or in a short period. This is different from the parking occupancy rate that is represented by percentages, namely the occupancy divided by the space capacity. In many curbside parking locations where no space markings are painted, due to various vehicle sizes, it might not have a precise figure as the capacity. This can be especially misleading for small blocks with only a few parking spaces. Here, we estimate the parking occupancy only to keep away the possible error from counting capacities. Without the loss of generality, one can simply convert the occupancy to occupancy rate if the capacity can be precisely determined.

We denote the actual parking occupancy at time t by $N(t)$, as discussed above, it is interpreted as the total number of parked vehicles at time t . We let $Z_i(t) \in \{0, 1\}$ denote whether the vehicle of the i -th transaction departs by time t , where 1 indicates being present and 0 being absent. Then, $N(t)$ is the sum of $Z_i(t)$ over all transactions plus the number of unpaid vehicles $U(t)$,

$$N(t) = \left(\sum_i Z_i(t) \right) + U(t) \quad (5)$$

To minimize the occupancy estimation error defined by a squared error loss function, the best estimate for $N(t)$ is its expectation $\mathbb{E}(N(t))$, which equals to the sum of expectations of the two terms.

$$\mathbb{E}(N(t)) = \mathbb{E}\left(\sum_i Z_i(t)\right) + \mathbb{E}(U(t)) \quad (6)$$

The first term $\mathbb{E}(\sum_i Z_i(t))$ can be obtained by aggregating $P(t)$ of all transactions at time t as well as the expected value of unpaid parking vehicles. If transaction i occurs after time t , namely $\mu_i - l_i < t$, then the parker associated with this transaction does not contribute to the parking occupancy. $\mathbb{E}(N(t))$ then reads,

$$\mathbb{E}(N(t)) = \sum_{i \in \{i \mid \mu_i - l_i \leq t\}} \mathbb{E}(Z_i(t)) + \mathbb{E}(U(t)) \quad (7)$$

$$= \sum_{i \in \{i \mid \mu_i - l_i \leq t\}} \frac{1}{1 + e^{\frac{t - \mu_i - a l_i - b}{c l_i + d}}} + \mathbb{E}(U(t)) \quad (8)$$

For the term of unpaid parking, $\mathbb{E}(U(t))$, the ground truth information regarding unpaid parking is lacking. We initially model it as a fixed portion of the total occupancy, that is $\mathbb{E}(U(t)) = r \cdot \mathbb{E}(N(t))$, where r is a new parameter in the model. During our experiments on the occupancy estimation, however, we found that due to the high level of dispersion of individuals' actual parking payment behavior and the excessive number of parameters to learn, the model does not converge very well in some cases, and it tends to overfit the data. To improve the performance of estimation while keeping the payment model reasonable, we do not explicitly consider unpaid parking except for those overnight unpaid parking in the morning (by Assumption 5 and addressed in Section 2.3). Rather, unpaid parking during the day is implicitly represented through the tendency towards underpay in the model. Unpaid parking results in a biased estimate of the probability of individuals' departure, namely individuals are estimated more likely to depart later than they actually are. Such a simplification of the model

is practically plausible for both occupancy estimation and patrolling guidance for enforcement officers.

By applying this simplification, Eq 7 is changed to

$$\mathbb{E}(N(t)) = \sum_{i \in \{i | \mu_i - l_i \leq t\}} \frac{1}{1 + e^{\frac{t - \mu_i - a l_i - b}{c l_i + d}}} \quad (9)$$

When the time interval is relatively small, e.g., 10 minutes, it is reasonable to assume the variation of occupancy rate within each 10-min interval is negligible, in other words, a constant occupancy rate can be observed throughout the 10-min interval. In the case when the time interval has to be large, such as the occupancy ground truth collected by *SFPark* sensors are only available on an hourly basis, it is not realistic to assume a constant occupancy during this entire hour. As a result, the occupancy estimate should be such that the probability of staying, $P(t)$, is integrated within the hour (60 minutes) and accumulated among all vehicles.

$$\begin{aligned} \mathbb{E}(N_{60}(t)) &= \sum_{i \in \{i | \mu_i - l_i \leq t\}} \int_t^{t+60} \mathbb{E}(Z_i(\tau)) d\tau \\ &= \sum_{i \in \{i | \mu_i - l_i \leq t\}} \int_t^{t+60} \left(\frac{1}{1 + e^{\frac{\tau - \mu_i - a l_i - b}{c l_i + d}}} \right) d\tau \\ &= \sum_{i \in \{i | \mu_i - l_i \leq t\}} F(\tau | \mu_i, l_i) \Big|_{\max(t, \mu_i - l_i)}^{t+60} \\ &= \sum_{i \in \{i | \mu_i - l_i \leq t\}} \left(\tau - (c \cdot l_i + d) \cdot \ln(e^{\frac{\tau - \mu_i}{c l_i + d}} + e^{\frac{a l_i + b}{c l_i + d}}) \right) \Big|_{\max(t, \mu_i - l_i)}^{t+60} \end{aligned} \quad (10)$$

Where $N_{60}(t)$ in Eq 10 denotes the total occupied minutes within 60 minutes starting at t , the overall occupancy rate for an area in that hour can be calculated by dividing $N_{60}(t)$ by the total minutes available for parking among all parking spots.

We denote the occupancy data collected by in-ground sensors or manual counts by $Y(t)$. The measurements of occupancy have errors due to sensors,

$$Y(t) = N(t) + \epsilon_{sensor} \quad (11)$$

where ϵ_{sensor} is the error term representing the sensor error and having zero mean.

When both meter transactions data and occupancy ground truth are available, the parameter set $\omega = (a, b, c, d)$ can be learned through minimizing the squared estimation error over all time intervals.

$$\min_{\omega} \sum_{t \in all} \left(Y(t) - \mathbb{E}(N(t, \omega)) \right)^2 \quad (12)$$

We will discuss an efficient algorithm to solve for the optimal ω in the next section.

In this paper, the performance of occupancy estimations is evaluated via Normalized Root Mean Squared Error (NRMSE) or Normalized Root Mean Squared Deviation (NRMSD), as

defined below,

$$NRMSE(NRMSD) = \sqrt{\frac{1}{M} \sum_{t \in all} \left(\frac{Y(t) - \mathbb{E}(N(t))}{Y(t)} \right)^2} \quad (13)$$

where M denotes the total number of time intervals. The NRMSE can be expressed as the percentage error of occupancy estimation.

2.3. Overnight parking model

As will be seen in our experiments, for some specific areas, overnight parked vehicles are of significant portion during AM hours. For instance, at 9:00 AM on Fewer Street, Pittsburgh, around 50% of all parked vehicles are overnight parking without timely payment. In another example, Civic Center areas of San Francisco, overnight parked vehicles leave gradually during AM hours. The following model is used to simulate their departure time by Assumption 5.

$$\mathbb{E}(R(t)) = \begin{cases} R_o \cdot \frac{1}{1 + \xi \cdot (t - t_s)} & t \leq t_E \\ 0 & t > t_E \end{cases} \quad (14)$$

Where R_o and ξ denote the total number of vehicles parked overnight and the departure rate of overnight parking vehicles respectfully, t_s is the enforcement starting time in early morning, and t_E is the end of overnight parking departure period. t is the time of day measured in minutes starts from the enforcement starting time (e.g., 9:00 AM). Let $\sigma = (R_o, \xi)$ be the parameter set of the overnight parking model. To avoid overfitting, the payment behavior model ω is first learned by analyzing PM hour data where no overnight parked vehicles are assumed. Then, the overnight parking model is learned given the payment model and the AM data,

$$\min_{\sigma} \sum_{t \in AM} \left(Y(t) - \mathbb{E}(N(t, \omega)) - \mathbb{E}(R(t, \sigma)) \right)^2 \quad (15)$$

According to our observations in certain areas in Pittsburgh, as will be seen from the experiments, the amount of overnight parked vehicles does not seem to decline gradually over time. Rather, they vanished quickly immediately after the enforcement starting time. Therefore, applying a simple model where a fixed portion of overnight parked vehicles depart within the same time interval can be a practical solution instead of a complicated model. The fixed portion can be obtained by minimizing the squared estimation error. In our experiments, we use either overnight parking model depending on our experience or goodness of fit to the available data.

2.4. Naive estimation

To evaluate the effectiveness of the parking payment models in terms of occupancy estimation accuracy, we use a naive estimation method as the baseline. The naive method is defined under a straightforward assumption that each vehicle departs exactly at the expiration time of the transaction, in other words, with no unpaid parking, no overpayment nor underpayment. The occupancy of an area at time t , using the naive method, can be estimated by counting the number of effective transactions that start before t and expire after t . By comparing NRMSEs of occupancy estimations between our model and the naive model, the gain of using probabilistic distributions for behavioral modeling can be evaluated.

2.5. Solution algorithm for parameter estimation

We adopt the gradient descent method to solve the optimization problem 12. Gradient descent is a first-order optimization algorithm that finds a local minimum of a defined and differentiable function. In each iteration, the algorithm takes one step proportional to the negative gradient of the objective function (Boyd & Vandenberghe 2004), as stated below.

$$\omega_{n+1} = \omega_n - \gamma_n \nabla F(\omega_n), \quad n \geq 0. \quad (16)$$

Where $F(X)$ is the objective function to minimize, ω_i is the set of parameters in the i -th iteration, and γ_i is a step size of the gradient descent in the i -th iteration. For the optimization problem 12, $F(X)$ is replaced by a regularized loss function $\ell(\omega)$,

$$\ell(\omega) = \frac{1}{M} \sum_{t \in all} \left(Y(t) - \mathbb{E}(N(t, \omega)) \right)^2 + \alpha \cdot \|\omega\|_2 \quad (17)$$

The first term measures the distance between the estimated occupancy $\mathbb{E}(N(t, \omega))$ and the observations $Y(t)$ at time t . The second term is a regularization term penalizing the complexity of the model through an L2 norm, a way of preventing overfitting. The hyperparameter α for the regularization term is set to 0.2 in this study.

For this gradient descent solution algorithm, we adopt a fixed step size $\gamma_n = 0.005$, for all n . The termination criteria are that the overall squared estimation error becomes less than 0.5%, or the total number of iterations reaches 5000, whichever comes first.

Similarly, the overnight parking model is also trained by the gradient descend algorithm for the optimization problem 15.

3. Data

In this paper, we use two sets of data for model validation and behavioral insights. The two data sets are from areas with possibly very different parking and payment behavior, and thus our models can be fully examined.

3.1. On-street parking near Carnegie Mellon University campus

We possess all transactions of on-street parking meters near Carnegie Mellon University area in Pittsburgh on two Mondays: 21st (365 relevant records), and 28th (418 relevant records) of Oct 2014. Each transaction entry comes with the start and ending time, the meter location and the parking rate. The transaction data is acquired from Pittsburgh Parking Authority.

Also available on these two days are the number of on-street parked vehicles on the 10-min basis during peak hours (8 AM-10 AM and 3 PM-5 PM) on Frew Street, Tech Street, and Margaret Morrison St near Carnegie Mellon University campus, as well as the total number of parking spots for those blocks. This part of data was collected by manual counts. Those on-street parking spaces are enforced from 8 AM to 6 PM.

3.2. *SFPark* pilot areas

We also collected transactions of on-street parking meters for multiple *SFPark* pilot areas during March-April 2013, where smart meters were installed and meter prices were subjected to adjustment based on average occupancies of the previous months. All the transactions are provided by SFMTA through the website of *SFPark*¹. Each transaction entry comes with the starting and ending time, the meter location and the total price paid for this transaction.

The same website also provides on-street parking sensor data (SFMTA 2014). Occupancies are available at the street block level on an hourly basis. The occupancy data were collected and provided by SFMTA as part of the *SFPark* project. Areas of the sensor data are in correspondence to those of the transactions data. Each data point specifies the date, hour, road and block name as well as the total number of parking spaces on that blocks. It also provides the total occupied seconds, total vacant seconds and total unknown seconds (when status cannot be determined to be vacant or occupied, less than 5 percent of all data in general) on both sides of the street. Due to the limit of the stability and reliability of in-ground sensors, the accuracy of occupancy data is generally around 85% to 90% as tested by SFMTA.

4. Results and Analysis

In this section, the proposed method for occupancy estimation is applied to the two data sets described above. Several experiments are conducted to evaluate the performance of the method under various conditions. The payment probabilistic model is learned to reveal the parking payment behavior.

4.1. *CMU* parking area

4.1.1. *Occupancy estimation*

The occupancy ground truth information for the *CMU* parking area was manually counted during AM and PM peak hours (8 AM-10 AM; 3 PM-5 PM) on the 21st and 28th (Mondays), of October, 2014. We utilize all data of Oct. 21st as the training set and Oct. 28th as the testing set.

First, a naive estimation of occupancy using the transaction data on Oct 21 is performed. The result is shown in the top two figures of Fig 2, in which the x-axis stands for the time of day, and y-axis is the parking occupancy. As shown in the curves in figures, for PM hours, the naive estimation (dashed red curve) is mostly coherent with manual count (solid blue curve). This indicates that during the PM hours, most parkers follow their payment fairly closely, very little evidence of over or underpayment. For AM hours, however, there is a significant gap between the naive estimation and actual counts. A significant amount of vehicles park without paying or underpaying, which may be explained by those overnight parked vehicles since parking is free until 8 AM and parking enforcement is relatively loose in the early morning. Although it can be inferred that overnight parking has vanished by the PM period, it is difficult to know the exact departure time of those vehicles due to the lack of ground truth data from 10 AM to 3 PM. Based on our assumption and discussion regarding overnight parking in Section 2.1.3, we first

¹Data retrieved on <http://sfpark.org/about-the-project/pilot-evaluation/>

learn the probabilistic payment model on PM peak without applying overnight parking model, then we learn the overnight parking model separately via data of the AM peak.

We apply the gradient descent algorithm to solve for the optimization model 12 using the PM data. The optimal parameters are found as follow: $\omega = (a, b, c, d) = (0.080, -0.128, 0.14, -0.05)$

$$P(t|\mu, l) = \frac{1}{1 + e^{\frac{t - \mu - 0.08l + 0.128}{0.14l - 0.05}}} \quad (18)$$

Where l, t, μ are all measured in minutes, and $\omega = (a, b, c, d)$ is a dimensionless scalar vector.

$P(t|\mu, l)$ provides the probability of a vehicle who pays for duration l until expiration time μ and still parking at time t , as plotted in Fig 3. The red and blue curves are $P(t)$ of 120- and 60-min paid durations, respectively.

As shown in the middle left sub-figure of Fig 2, the training error of the payment model for PM hours is 2.2%, revealing an improvement comparing to the 3.5% from the naive estimation. After the PM model is obtained, the overnight parking is then added to the AM model by applying a constant portion of all parked vehicles. Occupancies are estimated for all AM intervals (i.e., 10 minutes). The final training error is shown in the bottom left sub-figure of Fig 2, the overall training NRMSE for both AM and PM hours is 3.42%.

Place Fig 2 Here.

Place Fig 3 Here.

Next, we test our payment model using the data set of Oct. 28th. The results are shown in Fig 2. In comparison to the naive estimation, our model improves the estimation as the NRMSE drops from 6.59% to 4.74% for PM hours. After overnight parking compensation is applied, the overall NRMSE of estimation from the payment model is 5.2% comparing to 12.16% from the naive estimation.

4.1.2. Payment behavior

The curves in Fig 3 show that the probability of a vehicle parking beyond the expiration time is around 0.6, implying an overall tendency of underpaid parking near CMU area. Throughout the day, 95% of parkers who paid for a 60-min parking depart within 40 min beyond their expiration time, and 80% of them depart within 20 min beyond expiration. Whereas for parkers who paid for a much longer parking session, say, 120 min, the compliance to the payment is generally lower. 77% (62%) of parkers depart within 40 (20) min beyond their expiration time. This roughly quantifies the variation of parkers' compliance to the payment and can help enforcement officers estimate the chance of parking violation.

According to the model formulation, parameter a is controlling the trend of payment brought by the length of parking durations, in this case of CMU area, a being positive implies that the chance of underpaying, or encountering parking violations will be higher for those paying for longer durations. The parameter c is determining the behavioral volatilities caused by the length of parking durations. In general, a longer parking duration brings more uncertainty and randomness towards parker's parking payment behavior.

4.1.3. Effective granularity for occupancy estimation

We have estimated the time-dependent occupancy for the entire surrounding on-street parking area near CMU, in all 150 parking spots. The estimation seems satisfactory at this gran-

ularity. Both parkers and parking managers would prefer the highest possible geographic resolution when estimating parking occupancies. However, as the spatial-temporal granularity of the estimation gets finer, the variation of parking payment behavior becomes larger, which can overwhelm the estimation. We thus examine the effect of estimation spatio-temporal granularity on the estimation error, and attempt to find the effective granularity balancing resolution and accuracy.

Place Fig 4 here.

As shown in Fig 4, the same payment model is applied to each of the four road blocks near CMU campus. Naturally, testing errors for each block are higher comparing to that of the entire area. Consistent with previous results, our payment model outperforms the naive estimation in each of the four blocks. Particularly, the estimation error of West side of Margaret Morrison Street is significantly higher than other blocks. In other words, the tuned payment model is either unable to capture the payment behavior on this date or over-fitted for a small block like this (with around 15 parking spots). For instance, people may park on a block and pay on parking kiosks at nearby blocks, or it could be that more people do not pay for parking on this date than other dates. The uncertainty and behavioral randomness are much higher in smaller blocks than in larger areas.

For occupancy estimation on a block with less than 20 total parking stalls, the NRMSE is significantly higher than a block with more than 30 stall numbers. For the purpose of disseminating information and dynamic pricing, a tolerance threshold can be set up for obtaining the minimum spatial resolution for occupancy estimation. In this case, if 10% serves the tolerance threshold, then we should provide occupancy estimations for three blocks, Frew Street West, Tech Street, and a combined block of Frew Street East and Margaret Morrison West.

4.2. *SFPark pilot areas*

After evaluating the estimation performance in a small parking area where traffic patterns are likely to be repetitive from day to day, now we turn to a much greater parking area in San Francisco (SF) where the traffic/parking patterns may be more random.

4.2.1. *Occupancy estimation*

The occupancy ground truth information of *SFPark* pilot project is collected by in-ground parking sensors, which have been reported to be malfunctioning or damaged from time to time. In order to alleviate the impact of such sensor failures in the study, only dates and blocks which have at least 23 hours of valid sensor records per day are picked for the experiment. Unlike the experiments on CMU data, parking occupancies for SF are estimated on an hourly basis, which is the highest temporal resolution available from sensor data. Finally, we use those sensor data and the corresponding meter transactions near Civic Center areas on five Fridays (namely 8th, 15th and 29th of March 2013, 5th and 19th of April 2013). A time-dependent payment probabilistic model is learned from those five Fridays. To evaluate the performance of the payment model, a five-fold cross-validation is executed. For each fold, data from one of the five Fridays are treated as a testing set, while the rest four are used as the training set. In Fig 5 and Fig 6, dashed red curves represent the parking occupancy estimated via naive or the proposed method; The solid blue curve stands for the occupancy ground truth; And the dotted purple curves are the 95% confidence intervals of the estimation.

We first compare the results of the naive estimation to the ground truth for all selected days. This provides a general idea of whether or not overnight parking, overpaid and underpaid parking are significant. It is found that the payment behavior near the Civic Center area vary substantially by time of day. As shown in the left figure of Fig 5, overnight unpaid parking exists. It is assumed that overnight unpaid vehicles gradually leave prior to 11 AM in this model. Thus, the overnight parking model is applied to 9 AM-11 AM only. There is a general tendency of people overpaying for parking during late morning and noon (11 AM-2 PM), since the naive estimation yields more occupancies than sensor measures. Travelers start to underpay during late afternoons hours. Particularly, at the end of pay parking period (5 PM-6 PM), the number of underpaid parkers is increasingly significant. A time-dependent payment probabilistic model for SF Civic area is adopted by dividing the time of day into four time periods: 9 AM-11 AM; 11 AM-2 PM; 2 PM-5 PM; 5 PM-6 PM. Parameters for the model are trained independently for each period.

Place Fig 5 here.

The final model for Fridays of Civic Center area is described as below. Training results are plotted in Fig 5, with the overall NRMSE 2.21%.

- 9 AM-11 AM: Apply a probabilistic payment model with parameter set 1: $\omega = (-0.054, -2.43, 0.17, 0.60)$, and overnight parking model with $\sigma = (1.55, 0.0082)$;
- 11 AM-2 PM: Apply a probabilistic payment model with parameter set 1: $\omega = (-0.054, -2.43, 0.17, 0.60)$.
- 2 PM-5 PM: Apply a probabilistic payment model with parameter set 2: $\omega = (0.012, 0.86, 0.21, 0.04)$.
- 5 PM-6 PM: Apply a probabilistic payment model with parameter set 3: $\omega = (0.049, 9.39, 0.29, 0.12)$.

Place Fig 6 here.

The testing results of the cross-validations, as well as the corresponding naive estimations are shown in Fig 6. On the right plots, the true occupancies (solid blue curve) are fairly close to the model estimations (dashed red curve) and fall in the 95% confidence intervals represented by the dotted purple curves in most cases. As indicated by the NRMSEs, the proposed payment probabilistic model outperforms the naive estimation in all five cases. In terms of the overall accuracy of the occupancy estimation, the combination of payment model and overnight parking model yields better performance in four out of five cases. In fact, using the naive estimation with around 12% error rate can lead to improper parking prices, following SFpark's target of 60-80% optimal occupancies (Shoup 1997).

Next, we apply both the naive estimation and the probabilistic payment model to each day of week (from Monday through Thursday). It is found that Monday through Thursday, Fridays, and weekends each exhibit a typical pattern of parking occupancies and overall payment behavior. Among those days from Monday through Thursday, the tuned parameters for the payment models are fairly close to each other. Therefore, we consolidate Monday through Thursday, and use one single set of payment models, summarized as follows,

- 9 AM-11 AM: Apply a probabilistic payment model with parameter set 1: $\omega = (0.032, 1.40, 0.27, 0.12)$, and an overnight parking model with $\sigma = (1.6, 0.0467)$;

- 11 AM-2 PM: Apply a probabilistic payment model with parameter set 1: $\omega = (0.032, 1.40, 0.27, 0.12)$.
- 2 PM-6 PM: Apply a probabilistic payment model with parameter set 2: $\omega = (0.027, 2.20, 0.19, 0.11)$.

The overall testing NRMSE of the payment model is 5.2%, which consistently outperforms the naive estimation with the NRMSE 13.6%.

4.2.2. Time-of-day payment behavior

With the calibrated payment models, we are able to quantify the overpay/underpay behavior for the Civic Center areas. We focus on Fridays as their results are more interesting than other weekdays.

Fig 7 plots the total number of sessions paid and their average duration, both against hour of day. This indicates that generally the arrival rate reaches the peak around noon till 2 PM near Civic Center and the demand goes down in the late afternoon, both of which are consistent with the occupancy ground truth. However, travelers pay or park for a much longer period of time in the morning than the afternoon. Interestingly, parkers around Civic center areas are more likely to overpay when their anticipated duration is longer, possibly due to higher uncertainty related to longer durations. When their anticipated duration is shorter, they are more likely to be risky and underpay.

Place Fig 7 here.

The three fitted probabilistic payment models (using paid duration 120 minutes as an example) are shown in Fig 8. On Friday mornings, vehicles on average leave 5 minutes before their sessions expire, whereas, in the afternoon (2 PM-5 PM), parkers leave 3 minutes later after their sessions expire, for the end of pay parking period (5 PM-6 PM), people on average leave 15 min later than their session expiration time. Moreover, from the calibrated parameters of the three models as well as the CMU model in Sec 4.1.2, it is found that the overall payment behavior is determined by a . A positive a indicates that, if generally underlying, people tend to underpay more for longer parking durations, whereas a negative a infers the opposite. The CMU model, as well as the two afternoon models of Civic Center area with positive a values, indicate underpaying behavior in general, where the morning model of Civic Center with a negative a shows overall overpaying behavior. On the other hand, as the parameter c of the afternoon model is greater than that of the morning and noon model in Civic Center area, the payment behavior is more volatile in the afternoon than in the morning. As a result, prediction in the morning is generally more accurate than in the afternoon.

Comparing this model with the one for CMU areas in Fig 3, the parameter c of the CMU model is less than that of the Civic Center model, implying the probability of parkers' punctual departure near the Civic Center is greater. For the same paid duration (e.g., 120 min), parking payment behavior is more uncertain or random for the Civic Center than CMU areas. Thus, the variation of the estimation error for parking occupancies near CMU area are generally less than the Civic Center area.

The results of parking payment behavior near the Civic Center implies that morning parkers generally anticipate a longer duration and tend to overpay in general. Afternoon parkers would otherwise anticipate a shorter duration and tend to risk not paying full. This is possibly due to the nature of parking trips and enforcement coverage. Morning trips are more likely to be business trips, and parking enforcement may be more intense. Many afternoon trips are

individual trips, and parkers tend to risk, especially as the time approaches the enforcement ending time.

Place Fig 8 here.

4.2.3. *Day-of-week payment behavior*

We plot the naive estimation results on the dataset of March 2013 for each typical day of week in Fig 9, namely Monday through Thursday, Fridays, and weekends. For weekdays, travelers generally tend to slight underpay for most the day, except for Friday late mornings when people tend to slight overpay. The slight underpay may be consistent with the nature of weekday trips of which a large portion are made by local residents. Parkers tend to know the areas very well, and are willing to take a small risk of underpaying for minutes. On weekends, parking spaces near the Civic Center are significantly overpaid throughout the whole day, which can be possibly explained by the nature of leisure trips made by tourists. Apparently, overnight parked vehicles may still occupy space in the early morning after the enforcement starting time, especially on weekdays. This can lead to significant revenue loss.

Place Fig 9 here.

4.2.4. *Payment behaviors at various locations*

As discussed before, parking near CMU (as a typical near-campus parking) reveals a relatively low variation of departure time from the expiration time, and generally, parkers do not tend to overpay nor underpay significantly. While for business districts such as the Civic Center of San Francisco, the departure time variance is much greater, and travelers are generally underpaying on weekdays and overpaying on weekends. In addition, it is also found that the payment behavior varies substantially by parking blocks within the Civic Center area.

Place Fig 10 here.

Figure 10 shows two typical occupancy patterns on two example blocks near the Civic Center. On the Oak St, the naive estimation consistently under-estimates the parking occupancy throughout the day, implying a clear trend of underpaying or unpaid parking. To the opposite, for Golden Gate Ave., there is a strong evidence of overpaid parking. Based on such a discrepancy, we believe that the parking payment behavior may be related to the duration of purchased parking sessions, caused by the nature of those trips. As indicated by the bottom left figure of Fig 10, the average duration of paid sessions on the Oak St. is generally shorter than that on Golden Gate Ave. In other words, for crowded downtown locations like San Francisco Civic Center, people tend to underpay for short-term parking than for long-term. Possible explanations could be that some of the parkers assuming the risk of getting parking tickets are low when they are parking for only a short period of time. While people usually being conservative when estimating their departure time for relatively longer parking sessions, which results in an overpaying pattern in general.

4.2.5. *Effective granularities for occupancy estimation*

Previous subsections have shown that the probabilistic payment model results in a good performance on estimation for the entire Civic Center area on an hourly basis, which is a relatively large spatio-temporal granularity, around 500 parking spots per hourly estimation. To explore a

proper granularity with satisfactory accuracy as well as resolution, we examine various spatio-temporal granularities for estimating parking occupancies near the Civic Center area.

Place Fig 11 here.

In order to evaluate the performance of the proposed model under various spatial-temporal granularities, twelve parking zones are chosen. Their total number of parking spots ranges from 9 to 267. Parameters of the payment model, as well as overnight parking model, are estimated for each zone individually on Fridays of March 2013. The models are then tested by the data on Fridays of April 2013. The results of model testing are shown in Fig 11.

The plots in Fig 11 show that when estimating occupancy for a small block with 9, 13, and 26 total spots, the performance of the estimation is rather poor. The NRMSE generally goes down as the size of parking spaces increases. The estimation on those testing datasets is the most satisfactory when the number of total parking spots is more than 174. Similar to the discussion on CMU parking area, for the purpose of disseminating estimated occupancy information and dynamic parking pricing, we set the tolerance threshold as 10% again. Thus, estimating occupancy in a total of 150-200 spaces on an hourly base estimation sounds a reasonable granularity.

By comparing the results of granularities test between SF Civic Center area and CMU parking area, we can see that the most cost-effective granularity for occupancy estimation varies by location. CMU parking area exhibits a much lower variation in payment behavior than the Civic Center area. Specifically, most of the people parking in CMU area are university faculty, staff, and students, whose behaviors towards daily parking payment are in general repetitive from day to day. Most of them have fixed and recurrent travel plan every day and thus usually choose to park in similar locations, which makes their daily parking schedule stable and easy to be predicted. This results in a higher estimation granularity. On the other hand, for urban areas and business districts like Civic Center area, the compositions of parking population are complex: tourists, short time visitors, long-term daily parkers, all of which come with different parking habits and behavior. In addition, temporal unpaid parking such as loading trucks also occupied a fairly large amount of parking resources, which are usually difficult to monitor via transactions. As a result, the effective granularity for reliable estimation is generally lower.

5. Conclusions

In this study, we propose to estimate time-varying parking occupancy from parking ticket transaction data by considering probabilistic on-street parking payment behavior. The payment model can be used to analyze various aspects of parkers' behavior including the tendency of underpaying or overpaying, overnight unpaid parking, and the variation of their departure time from their purchased expiration time. As indicated by the experiments (CMU area, and the Civic Center area in SF), the model is effective in capturing on-street parking payment behavior under various scenarios, and outperforms other on-street parking models in the literature, including the linear regression method used by *SFPark* as well as the naive model. It generally provides reliable estimations of time-varying occupancies under relatively low spatial-temporal granularities. This model quantifies the variation of parkers' compliance to the payment and can help enforcement officers estimate the frequencies of parking violations in certain location and time period.

The main findings of this paper are, 1) the parking ticket transactions can be effectively translated into useful time-varying occupancy information. This provides a cheap alternative solution to help deploy parking management and guidance system, without necessarily installing parking sensors. The estimation is generally satisfactory. When the probabilistic payment model is applied to areas with stabilized parking patterns and to areas with random parking patterns, the estimation error for the former case is much smaller than the later one; 2) people tend to underpay in CMU area in general, whereas for Civic Center area, payment behavior varies for different days of week as well as time of day. For Fridays, people tend to overpay in the mornings and their overall behavior changes to underpay during afternoon hours. Also, parkers' payment behavior, in general, is more random in Civic Center area than in CMU areas; 3) there exists an effective granularity for occupancy estimation, defined as the highest spatial resolution for this model to perform reliably. For CMU area, it is around 10-20 spaces for each street-block, while the effective granularity is 150-200 spaces for the Civic Center area due to more random parking behavior.

According to our analysis for the SFpark area, we have two general suggestions. Firstly, temporal resolution of parking pricing can be further refined. One of the main goals of *SFPark* project is to maintain parking occupancy within a reasonable range. To fulfill this goal, meter rates are subjected to regular adjustments and different rates are applied to different time periods of a day. The current time periods are 9 AM to noon; noon to 3 PM and 3 PM to 6 PM. Both the ground truth and naive estimation have indicated that parking occupancies of Civic Center reach the peak within a day during 11 AM to 2 PM, see Fig 5. It is also found that on Fridays, parkers tend to overpay during 9 AM to 2 PM, as opposed to underpay in later afternoons. We thus suggest that temporal break-down for pricing of Civic Center area be modified to: 9 AM-11 AM, 11 AM-2 PM and 2 PM-6 PM for Monday to Thursday; 9 AM-11 AM, 11 AM-2 PM, 2 PM-5 PM and 5 PM-6 PM for Friday. Secondly, the frequencies of parking enforcement patrols can be decided according to the specific characteristics of parking payment behavior. As our study has discovered, people's parking payment behavior varies by different time and locations. In order to make the best use of parking enforcement resources, the patrolling frequencies for areas and time periods with significant observations of underpaying should be set higher than other areas.

Although the proposed method yields satisfying results in most cases, it is also limited by its nature of relying on information provided by parking transaction data and the assumption of homogeneous parking payment behavior among users. The model may not be suitable for places with an excessive amount of unpaid parking and extraordinary uncertainties in payment behaviors.

Place Fig 12 here.

The two figures of Fig 12 are results of naive estimation for general downtown SF for Monday through Thursday and Fridays respectively during October 2013. As indicated by the significant gaps between the curves, parking occupancies collected by sensors are more than twice as much as the naive estimation from transactions. It may be unrealistic to attribute such discrimination, which is almost half volume of the total occupancy, solely to underpaying parking. In fact, most of the parking stalls with sensors in this area are located on business or shopping districts, which are usually very busy during the daytime. Moreover, loading and standing trucks are frequent on these streets. Short-term unpaid parking and truck loading could be the actual causes of the gap between actual occupancies and transactions.

This paper has some limitations and can be improved in the future research. First, the probabilistic payment model can be improved by learning from parking enforcement data (such as citations) and truck loading data. These data provide partial information regarding how much parking space is occupied without being revealed in the transactions system. They can be incorporated to improve the model accuracy. Second, this study builds a general model that takes the transactions of parking on those traditional kiosks or model applications where users have to purchase or renew parking for at least a certain time interval (such as 15 minutes) without knowing the precise departure time. As payment/sensing technologies improve, we may know precisely how long a fraction of parkers stayed. For instance, payment upon tap-in and tap-out of a mobile application is offered to precisely record the parking departure time, so users can pay for the actual period of time parked. This information would be helpful in reducing the variance of the overall occupancy estimation and improve accuracy. Third, since the occupancy observations are necessary to tune the model but are quite limited due to budget, it is worthwhile to examine the value of the ground truth information that varies by location and time. It would be interesting to optimally allocate the sensing resources of collecting ground truth observations in the parking network, in order to estimate occupancy for all blocks most effectively. Likewise, we would need to estimate the budget size of ground truth collection for the proposed model to be applicable and accurate for all blocks. Last but not least, based on the payment behavior learned in this study, we can recommend the parking agencies the optimal patrolling strategies and routes for the purpose of enforcement.

Acknowledgement

This research is funded by National Science Foundation Award CNS-1544826, and in part by Carnegie Mellon University's Technologies for Safe and Efficient Transportation, a National USDOT University Transportation Center for Safety (T-SET UTC) sponsored by the US Department of Transportation. The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. The U.S. Government assumes no liability for the contents or use thereof. The authors would like to thank

the Pittsburgh Parking Authority for providing parking data.

References

- Arnott, R. (2006), 'Spatial competition between parking garages and downtown parking policy', *Transport Policy* **13**(6), 458–469.
- Arnott, R. & Inci, E. (2006), 'An integrated model of downtown parking and traffic congestion', *Journal of Urban Economics* **60**(3), 418–442.
- Balac, M., Ciari, F. & Axhausen, K. W. (2017), 'Modeling the impact of parking price policy on free-floating carsharing: Case study for zurich, switzerland', *Transportation Research Part C: Emerging Technologies* **77**, 207–225.
- Boyd, S. & Vandenberghe, L. (2004), *Convex optimization*, Cambridge university press.
- Cao, J. & Menendez, M. (2013), 'Methodology to evaluate cost and accuracy of parking patrol surveys', *Transportation Research Record: Journal of the Transportation Research Board* (2359), 1–9.
- Chen, M., Hu, C. & Chang, T. (2011), The research on optimal parking space choice model in parking lots, in 'Computer Research and Development (ICCRD), 2011 3rd International Conference on', Vol. 2, IEEE, pp. 93–97.
- Chen, Z., Xia, J. C., Irawan, B. & Caulfield, C. (2014), 'Development of location-based services for recommending departure stations to park and ride users', *Transportation Research Part C: Emerging Technologies* **48**, 256–268.
- Cleveland, D. (1963), 'Accuracy of the periodic check parking study', *Traffic Engineering* **33**(12), 14–17.
- Cullinane, K. & Polak, J. (1992), 'Illegal parking and the enforcement of parking regulations: causes, effects and interactions', *Transport Reviews* **12**(1), 49–75.
- Hsu, C.-H. (1989), 'Electronic parking sensor apparatus'. US Patent 4,855,736.
- Lei, C. & Ouyang, Y. (2017), 'Dynamic pricing and reservation for intelligent urban parking management', *Transportation Research Part C: Emerging Technologies* **77**, 226–244.
- Mathur, S., Jin, T., Kasturirangan, N., Chandrasekaran, J., Xue, W., Gruteser, M. & Trappe, W. (2010), Parknet: drive-by sensing of road-side parking statistics, in 'Proceedings of the 8th international conference on Mobile systems, applications, and services', ACM, pp. 123–136.
- Millard-Ball, A., Weinberger, R. R. & Hampshire, R. C. (2014), 'Is the curb 80% full or 20% empty? assessing the impacts of san franciscos parking pricing experiment', *Transportation Research Part A: Policy and Practice* **63**, 76–92.
- Nawaz, S., Efstratiou, C. & Mascolo, C. (2013), Parksense: A smartphone based sensing system for on-street parking, in 'Proceedings of the 19th annual international conference on Mobile computing & networking', ACM, pp. 75–86.

- 1
- 2
- 3
- 4
- 5
- 6 Ottosson, D. B., Chen, C., Wang, T. & Lin, H. (2013), 'The sensitivity of on-street parking
- 7 demand in response to price changes: A case study in seattle, wa', *Transport Policy* **25**, 222–
- 8 232.
- 9
- 10 Outlook, T. (2011), Meeting the needs of 9 billion people, in 'International Transport Fo-
- 11 rum/OECD, ITF', pp. 2–11.
- 12
- 13
- 14 Pierce, G. & Shoup, D. (2013), 'Getting the prices right: an evaluation of pricing parking by
- 15 demand in san francisco', *Journal of the American Planning Association* **79**(1), 67–81.
- 16
- 17 Pu, Z., Li, Z., Ash, J., Zhu, W. & Wang, Y. (2017), 'Evaluation of spatial heterogeneity in the
- 18 sensitivity of on-street parking occupancy to price change', *Transportation Research Part C: Emerging Technologies* **77**, 67–79.
- 19
- 20
- 21
- 22 Qian, Z. & Rajagopal, R. (2015), 'Optimal dynamic pricing for morning commute parking',
- 23 *Transportmetrica A: Transport Science* **11**(4), 291–316.
- 24
- 25 Qian, Z. S. & Rajagopal, R. (2014), 'Optimal dynamic parking pricing for morning commute
- 26 considering expected cruising time', *Transportation Research Part C: Emerging Technolo-*
- 27 *gies* **48**, 468–490.
- 28
- 29
- 30 Richardson, A. (1974), An improved parking duration study method, in 'Australian Road Re-
- 31 search Board (ARRB) Conference, 7th, 1974, Adelaide', Vol. 7.
- 32
- 33 Rodier, C. J. & Shaheen, S. A. (2010), 'Transit-based smart parking: An evaluation of the
- 34 san francisco bay area field test', *Transportation Research Part C: Emerging Technologies*
- 35 **18**(2), 225–233.
- 36
- 37
- 38 SEATTLE, C. O. (2002), 'Seattle parking management study'.
- 39
- 40 Seattle, D. (2011a), Citywide paid parking study and 2011 rates, executive summary revised,
- 41 Technical report.
- 42
- 43 Seattle, D. (2011b), Citywide paid parking study. p. 17., Technical report.
- 44
- 45 SFMTA (2014), Parking sensor technology performance evaluation, Technical report.
- 46
- 47 Shoup, D. C. (1997), 'The high cost of free parking', *Journal of planning education and re-*
- 48 *search* **17**(1), 3–20.
- 49
- 50
- 51 Sousanis, J. (2011), 'World vehicle population tops 1 billion units'.
- 52
- 53 Verhoef, E., Nijkamp, P. & Rietveld, P. (1995), 'The economics of regulatory parking policies:
- 54 the (im) possibilities of parking policies in traffic regulation', *Transportation Research Part*
- 55 *A: Policy and Practice* **29**(2), 141–156.
- 56
- 57 Weinberger, R., Kaehny, J. & Rufo, M. (2010), 'Us parking policies: An overview of manage-
- 58 ment strategies'.
- 59
- 60
- 61
- 62
- 63
- 64
- 65

Abstract

Over 95% of on-street paid parking stalls are managed by parking meters or kiosks. By analyzing meter transactions data, this paper provides a methodology to estimate on-street time-varying parking occupancy and understand payment behavior in an effective and inexpensive way. We propose a probabilistic payment model to simulate individual payment and parking behavior for each parker. Aggregating the payment/parking of all transactions leads to time-varying occupancy estimation. Two data sets are used to evaluate the methodology, parking spaces near Carnegie Mellon University (CMU) campus, and near the Civic Center in San Francisco. The proposed model generally provides reliable estimations of occupancies at a low error rate and substantially outperforms other naive models in the literature. From the results of the experiments we find that people generally tend to slightly underpay in CMU area, whereas for Civic Center area, payment behavior varies by time of day and day of week. For Fridays, people generally tend to overpay and stay longer in the mornings compared to underpaying and parking for shorter durations in the late afternoons. Parkers' payment behavior, in general, is more variable and noisier around Civic Center than around CMU. Moreover, we explore the effective granularity, defined as the highest spatial resolution for this model to perform reliably. For CMU areas, the effective granularity is around 10-20 spaces for each block of streets, while it is 150-200 spaces for the Civic Center area due to more random parking behavior.

1
2
3
4
5
6 **Keywords**
7

8 Time-dependent parking occupancy; payment behavior; parking meter transactions
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Figures

Figure 1: Schematic diagram of the individual payment behavior model

Figure 2: Performance evaluation of the Payment Probabilistic Model on CMU parking area. Occupancies are measured by total occupied parking hours per hour.

Solid blue curve: Ground truth collected by manual count; Dashed red curve: Occupancy estimations by model; Dotted purple curve: Boundary of 95% confidence interval;

LeftTop: Naive estimation on data of Oct 21st.

LeftMiddle: Payment probabilistic model: Training result without overnight model.

LeftBottom: Payment probabilistic model: Training result with overnight mode.

RightTop: Naive estimation on data of Oct 28th.

RightMiddle: Payment probabilistic model: Testing result without overnight model.

RightBottom: Payment probabilistic model: Testing result with overnight model.

Figure 3: $P(t)$: individual payment model for parkers near CMU area. The x-axis is time in minutes with respect to the paid expiration time (negative indicates overpaying, positive indicates underpaying); y-axis is the probability of a vehicle still parking at time t . Red solid curve: 120 min paid duration; Blue dashed curve: 60 min paid duration.

Figure 4: Testing results of occupancy prediction (Oct 28th) for each of the four roadblocks.

LeftTop: West side of Frew Street: Estimated parking spaces: 65; NRMSE: 9.1%.

RightTop: Tech St : Estimated parking spaces: 33; NRMSE: 10.3%.

LeftBottom: East side of Frew Street: Estimated parking spaces: 17; NRMSE: 14.88%.

RightBottom: West side of Margaret Morrison Street: Estimated parking spaces: 15; NRMSE: 46.8%.

Figure 5: Training results of payment model on Fridays during March and April 2013 of Civic Center area. Occupancies are measured in total occupied parking hours per hour.

Left: Naive estimation. Overall NRMSE: 10.53%; 11AM-6PM NRMSE: 5.67%.

Right: Payment probabilistic model. Overall NRMSE: 1.48%; 11AM-6PM NRMSE: 1.66%.

Figure 6: Cross-validation on five Fridays of March-April 2013 of Civic Center Area. Results of model testings on five Fridays are plotted on the left, and the corresponding naive estimations are plotted on the right.

Figure 7: Civic Center area on Fridays: Vehicle arrival rate and the average purchased parking durations.

Figure 8: $P(t)$: individual payment models for parkers paying for 120-min parking of Civic Center area on Fridays. The x-axis is the time in minutes prior to the paid expiration time (negative indicates overpay, positive indicates underpay); y-axis is the probability of a vehicle still parking at time t . Red solid line: 9 AM-1 PM; Blue dashed line: 2 PM-5 PM; Purple dashed line: 5 PM-6 PM

Figure 9: Day-of-week payment behavior: Naive estimations of parking occupancy on March 2013.

Figure 10: Comparison of parking payment behaviors among blocks in Civic center area.

Topleft: Naive estimation of Oak St 0-100 on Fridays.

Bottomright: Parking arrival rate and average length of purchased session on Oak St 0-100 on Fridays.

Topright: Naive estimation of Golden Gate Ave 600-700 on Fridays.

Bottomright: Parking arrival rate and average length of purchased session on Golden Gate Ave 600-700 on Fridays.

Figure 11: Granularity evaluation of the payment behavior model: testing results of model fitting on fifteen individual parking blocks on Civic Center area.

Figure 12: Parking payment behaviors of general downtown SF

Left: Naive estimation on Monday - Thursday.

Right: Naive estimation on Fridays.

List of Figure captions

- Fig 1 Schematic diagram of the individual payment behavior model.
- Fig 2 Performance evaluation of the Payment Probabilistic Model on CMU parking area. Occupancies are measured by total occupied parking hours per hour. Solid blue curve: Ground truth collected by manual count; Dashed red curve: Occupancy estimations by model; Dotted purple curve: Boundary of 95% confidence interval;
LeftTop: Naive estimation on data of Oct 21st.
LeftMiddle: Payment probabilistic model: Training result without overnight model.
LeftBottom: Payment probabilistic model: Training result with overnight mode.
RightTop: Naive estimation on data of Oct 28th.
RightMiddle: Payment probabilistic model: Testing result without overnight model.
RightBottom: Payment probabilistic model: Testing result with overnight model.
- Fig 3 $P(t)$: individual payment model for parkers near CMU area. The x-axis is time in minutes with respect to the paid expiration time (negative indicates overpaying, positive indicates underpaying); y-axis is the probability of a vehicle still parking at time t . Red solid curve: 120 min paid duration; Blue dashed curve: 60 min paid duration.
- Fig 4 Testing results of occupancy prediction (Oct 28th) for each of the four roadblocks.
LeftTop: West side of Frew Street: Estimated parking spaces: 65; NRMSE: 9.1%.
RightTop: Tech St : Estimated parking spaces: 33; NRMSE: 10.3%.
LeftBottom: East side of Frew Street: Estimated parking spaces: 17; NRMSE: 14.88%.
RightBottom: West side of Margaret Morrison Street: Estimated parking spaces: 15; NRMSE: 46.8%.
- Fig 5 Training results of payment model on Fridays during March and April 2013 of Civic Center area. Occupancies are measured in total occupied parking hours per hour.
Left: Naive estimation. Overall NRMSE: 10.53%; 11AM-6PM NRMSE: 5.67%.
Right: Payment probabilistic model. Overall NRMSE: 1.48%; 11AM-6PM NRMSE: 1.66%.
- Fig 6 Cross-validation on five Fridays of March-April 2013 of Civic Center Area. Results of model testings on five Fridays are plotted on the left, and the corresponding naive estimations are plotted on the right.
- Fig 7 Civic Center area on Fridays: Vehicle arrival rate and the average purchased parking durations.
- Fig 8 $P(t)$: individual payment models for parkers paying for 120-min parking of Civic Center area on Fridays. The x-axis is the time in minutes prior to the paid expiration time (negative indicates overpay, positive indicates underpay); y-axis is the probability of a vehicle still parking at time t . Red solid line: 9 AM-1 PM; Blue dashed line: 2 PM-5 PM; Purple dashed line: 5 PM-6 PM
- Fig 9 Day-of-week payment behavior: Naive estimations of parking occupancy on March 2013.

- Fig 10 Comparison of parking payment behaviors among blocks in Civic center area.
Topleft: Naive estimation of Oak St 0-100 on Fridays.
Bottomright: Parking arrival rate and average length of purchased session on Oak St 0-100 on Fridays.
Topright: Naive estimation of Golden Gate Ave 600-700 on Fridays.
Bottomright: Parking arrival rate and average length of purchased session on Golden Gate Ave 600-700 on Fridays.
- Fig 11 Granularity evaluation of the payment behavior model: testing results of model fitting on fifteen individual parking blocks on Civic Center area.
- Fig 12 Parking payment behaviors of general downtown SF
Left: Naive estimation on Monday - Thursday.
Right: Naive estimation on Fridays.