Transportation Research Part B 112 (2018) 19–39

Contents lists available at ScienceDirect

Transportation Research Part B

journal homepage: www.elsevier.com/locate/trb

How likely am I to find parking? – A practical model-based framework for predicting parking availability

Jun Xiao[a], Yingyan Lou[a,*], Joshua Frisby[b,1]

[a] School of Sustainable Engineering and Built Environment, Arizona State University, 660 S. College Avenue, Tempe, AZ 85281, United States
[b] Mathematics & Computer Science Department, Glendale Community College, 6000 W. Olive Avenue, Glendale, AZ 85302, United States

ARTICLE INFO

Article history:
Received 21 February 2017
Revised 21 February 2018
Accepted 4 April 2018
Available online 21 April 2018

ABSTRACT

Parking availability information (or occupancy of parking facility) is highly valued by travelers, and is one of the most important inputs to many parking models. This paper proposes a model-based practical framework to predict future occupancy from historical occupancy data alone. The framework consists of two modules: estimation of model parameters, and occupancy prediction. At the core of the predictive framework, a queuing model is employed to describe the stochastic occupancy change of a parking facility. While the underlying queuing model can be any reasonable model, we demonstrate the framework with the well-established continuous-time Markov M\M\C\C queue in this paper. The possibility of adopting a different queuing model that can potentially incorporate the parking-searching process is also discussed. The parameter estimation module and the occupancy prediction module are both built on the underlying queuing model. To apply the estimation and prediction methods in real world, a few practical considerations are accounted for in the framework with methods to handle variations of arrival and departure patterns from day to day and within a day, including special events. The proposed framework and models are validated using both simulated and real data. Our San Francisco case studies demonstrate that the parameters estimated offline can lead to accurate predictions of parking facility occupancy both with and without real-time update. We also performed extensive numerical experiments to compare the proposed framework and methods with several pure machine-learning methods in recent literature. It is found that our approach delivers equal or better performance, but requires a computation time that is orders of magnitude less to tune and train the model. Additionally, our approach can predict for any time in the future with one training process, while pure machine-learning methods have to train a specific model for a different prediction interval to achieve the same level of accuracy.

1. Introduction

Searching for parking is a real struggle faced by many drivers, especially in urban areas (IBM, 2011). Empirical studies conducted in the US suggested that the average time spent searching for a curbside parking space ranged between 3.5 and 14 min (Shoup, 2006), which has tremendous negative impact on a traffic network. According to Shoup (2006), cruising for

* Corresponding author.
E-mail addresses: jun.xiao.1@asu.edu (J. Xiao), yingyan.lou@asu.edu (Y. Lou), joshua.frisby@asu.edu (J. Frisby).

[1] The work was performed when Mr. Frisby was at Arizona State University.

https://doi.org/10.1016/j.trb.2018.04.001

parking is responsible for 30% of traffic on average in downtown area. In a more recent research, Giuffrè et al. (2012) found that cruising for parking results in a peak increase of about 25–40% of the traffic flow. Moreover, congestion caused by cruising for parking is a waste of resources and aggravates environmental issues. Ayala et al. (2012) concluded that for a city like Chicago, 8.37 million gallons of gasoline would be consumed and 129 thousand tons of CO_2 emitted every year due to cruising for parking.

The significance of the parking searching problem has led to an increasing public demand for parking information and services recently. With the proliferation of advanced smartphones, a range of smartphone-based parking management services began to emerge. In fact, the International Parking Institute identified the prevalence of mobile applications the #2 emerging trend in parking in 2015 (IPI, 2015). Available services currently on the market include mobile payment (e.g., Pango 2016; ParkMobile 2016) and parking information provision and reservation (e.g., ParkMe Inc. 2016; Parkopedia 2016; ParkWhiz 2016; Spothero 2016; BestParking 2016). In addition to parking rates and facility properties provided by almost every parking application, ParkMe also offers real-time parking occupancy information (as a percentage) in selected markets.

On the other hand, the research communities have been studying various aspects of parking for decades. As early as in the 1950s, Vickrey suggested time-varying parking fees to ensure the economical utilization of parking spaces (Vickrey, 1994). To this day, policy and operational innovations aiming to improve social welfare is still a predominant topic in parking research. Such innovations include pricing (Arnott et al., 1991; Zhang and van Wee, 2011; Qian and Rajagopal, 2014; Mackowski et al., 2015; He et al., 2015), reservation (Teodorovic and Lucic, 2006; Boehlé et al., 2008; Delot et al., 2009; Caicedo et al., 2012; Liu et al., 2014; Shin and Jun, 2014; Chen et al., 2015), and information provision (Waterson et al., 2001; Thompson et al., 2001; Caicedo, 2009; Sun et al., 2016). Driven by the modeling need in these work, many studies have specifically investigated how parking competition affects both temporal and spatial travel patterns. The former is commonly achieved through bottleneck models (Yang et al., 2013; Xiao et al., 2016) while the latter through non-atomic games (more commonly referred to as traffic assignment) (Bifulco, 1993; Arnott and Rowse, 1999; Anderson and de Palma, 2004; Boyles et al., 2015) or more general game-theoretical framework (Ayala et al., 2011; Kokolaki et al., 2012; Guo et al., 2013). In addition, agent-based simulation employing rule-based or discrete choice models (Benenson et al., 2008; Waraich and Axhausen, 2012) has also been commonly adopted. It is worth noting that although information provision, more specifically the provision of parking occupancy and probability of finding a parking spot, is not as extensively studied as other parking management strategies, it has attracted more attention recently and is being considered by some of the models as key components, inputs, or assumptions (Caicedo et al., 2012; Shin and Jun, 2014; Qian and Rajagopal, 2013; Guo et al., 2013; Schlote et al., 2014; Boyles et al., 2015; Chen et al., 2015). Moreover, some recent empirical studies also found that parking occupancy information is mostly valued by drivers (Caicedo et al., 2006; van der Waerden et al., 2011) and such occupancy information (Caicedo et al., 2006) and related variables such as time spent searching for parking (Ibeas et al., 2014) and the availability of parking after a reasonable search time (Chaniotakis and Pel, 2015) do significantly affect drivers' parking choices.

In view of the significance of predictive parking occupancy information in parking searching, this paper aims to develop a model-based practical framework for parking occupancy prediction. The framework consists of two modules: estimation of model parameters, and occupancy prediction. At the core of the predictive framework, a queuing model is employed to describe the stochastic occupancy change of a parking facility. While the underlying queuing model can be any reasonable model, we demonstrate the framework with the well-established continuous-time Markov M\M\C\C queue in this paper. The parameter estimation module and the occupancy prediction module are both built on the underlying queuing model. The possibility of adopting a different queuing model that can potentially incorporate the parking-searching process is further discussed. To apply the estimation and prediction methods in real world, a few practical considerations are further discussed and methods proposed to handle more complex arrival and departure variations from day to day and within a day, including special events.

The framework and the methods are validated using both simulated and real-world data. Our San Francisco case studies demonstrate that the parameters estimated offline can lead to accurate predictions of parking facility occupancy both with and without real-time occupancy. Even when the underlying model may be inaccurate due to limited data availability (such as parking occupancy data during special event), a potential drawback of model-based approaches, the prediction module can be easily modified to offer prediction that is more accurate. We also performed extensive numerical experiments to compare the proposed framework and methods with several pure machine-learning methods published in recent literature. It is found that our approach delivers equal or better performance, but requires a computation time that is orders of magnitude less to tune and train the model. Additionally, our model can predict for any time in the future with one training process while pure machine learning methods have to train a separate model for each time window of interest to achieve accurate prediction. Moreover, by adopting a model-based approach over pure statistical and machine-learning methods, the framework could potentially incorporate the parking searching process (which can be reflected in the stochastic arrival process in the queuing model) in parking occupancy prediction. This would enable investigation on network traffic effects as a result of parking availability information provision and the subsequent possibly altered parking search behaviors using game-theoretical approaches.

To this end, the contributions of this paper include: 1) a model-based practical parking occupancy prediction framework grounded in the underlying stochastic queuing process; 2) the first to validate a model-based parking occupancy prediction framework using real data; and 3) insights on the applicability and suitability of the proposed model-based framework from comparison with pure machine learning parking occupancy prediction methods.

2. Literature review on parking availability prediction

Two approaches to parking availability prediction exist in the literature.

One approach starts with establishing some underlying model for the parking process; and parking availability prediction relies on the estimation of model parameters. The underlying model usually explicitly considers the stochastic arrival and departure processes of a parking facility. The arrival process is commonly assumed Poisson. As for the departure process, it is often assumed the parking duration follows negative exponential distribution (Millard-Ball et al., 2014; Portilla et al., 2009) or discretized Gamma distribution (Caicedo et al., 2012). Caliskan et al. (2007) adopts the classic continuous-time birth-and-death queueing model with Poisson arrival and exponential inter-departure time. In their simulation work, they assumed that the departure rate is given, and employed a maximum likelihood approach to estimate the arrival rate from simulated occupancy data. They further predicted future occupancy with the estimated arrival rate and the given departure rate with an autoregressive Gaussian process. Similarly, Lu et al. (2009) estimates the probability that a parking facility is completely full using the same model but assumes that advanced technologies are in place to provide data on arrival and departure rates directly. Boyles et al. (2015) incorporates the same continuous-time Markov model to specify the probabilistic parking availability in their network equilibrium analysis. Caicedo et al. (2012) proposed a centralized system for parking request allocation that has parking availability forecast as one of its components. The system computes future parking availability forecast based on requests allocation and simulated parking durations from discretized Gamma distribution. Existing studies in this category are mainly theoretical, and have at best validated their models using simulation. In addition, most of the aforementioned studies are not able to estimate parameters for the arrival and departure processes simultaneously, and require additional assumptions for prediction.

Another approach, instead of modeling the underlying parking process, applies statistical and machine learning methods to predict future occupancy directly from the observed data. These methods include simple regression (McGuiness and McNeil, 1991; Burns and Faurot, 1992), database system (Dunning 2006), chaotic time series analysis (Liu et al., 2010), multivariate spatiotemporal regression (Rajabioun and Ioannou, 2015), neural network (Vlahogianni et al., 2016; Ji et al., 2015; Ziat et al., 2016; Tiedemann et al., 2015), and clustering (Tiedemann et al., 2015; Tamrazian et al., 2015). One drawback of these models is the extensive tuning of the model structure. For example, in neural network approaches, model structure parameters include the number of input nodes, the number of input layers/hidden nodes, learning rate and momentum. Extensive tuning is required to determine these parameters before the actual learning of weights of each input and hidden node can take place. Moreover, while these methods are able to provide accurate prediction for a given time window and can be very useful in implementations of parking information systems, their value in exploring innovative parking policies and operational strategies is limited. This is because such statistical and machine learning algorithms are isolated from other aspects of the transportation system; and they are not set up to account for changes in human behaviors and traffic flow in the entire network, which are ultimately reflected in the stochastic arrival and departure processes.

The predictive framework and methods developed in our work fall into the first approach, where the underlying stochastic parking process is explicitly modeled. Different from previous work in this category, we have validated our approaches using both simulated and real data. Moreover, our approach is also able to estimate at the same time both arrival and departure rates based on historical occupancy data alone. While our predictive framework also employs regression techniques under certain conditions, the governing equation (model structure) is derived from the analytical properties of the underlying Markov queueing model and does not require tuning.

3. Prediction framework

The prediction framework consists of a parameter estimation module and an occupancy prediction module. At the core of the predictive framework, a queuing model is employed to describe the stochastic occupancy change of a parking facility. While the underlying queuing model can be any reasonable model, we demonstrate the framework with the well-established continuous-time Markov M\M\C\C queue in this paper. Mathematical properties of the M\M\C\C model, such as the time-dependent expectation, together with its asymptote, and the time-dependent variance of the parking facility occupancy are derived analytically under specific conditions. Applying the underlying queuing model, two approaches are developed to estimate the arrival and departure rates from historical occupancy data. The first approach makes use of the analytical properties of the occupancy probability distribution and employs curve-fitting techniques; and is suitable when the parking facility is under-saturated. When the parking facility is over-saturated, the second approach applies maximum likelihood or least squares estimation directly. To apply any of the two base estimation methods in real world, a few practical considerations are further discussed and methods proposed to handle arrival and departure patterns that are more complex. For example, due to day-to-day variations in the arrival and departure patterns, data clustering should be applied first to arrange historical daily occupancy time series into groups. When analyzing time-series data of occupancy within a 24-h timeframe, the time points that may mark a change in the underlying arrival and departure rates need to be identified; and an iterative process is proposed in this paper for this purpose. A simple prediction method based on the analytical properties of the occupancy probability distribution can be applied for normal parking traffic. On the other hand, a modified prediction method is proposed to offer more accuracy during special events, for which data availability might be limited and the underlying parking searching process (and thus the arrival and departure patterns) is not well understood. The prediction module can be engaged both with and without real-time update, and does not require additional data other than observed occupancy at the beginning of the day or in real time.

3.1. Stochastic queueing model for parking process

In this section, we first provide a brief introduction of a widely used (Caliskan et al., 2007; Lu et al., 2009; Boyles, 2015) continuous-time Markov model (M/M/C/C queueing model) for the stochastic parking process of a single parking facility. Time-dependent analytical properties of the model are then derived.

It is worth mentioning that the underlying queuing model can be any reasonable model. One possible alternative is to consider a discrete-time model where both the arrival and departure are described by a binomial process. This could potentially allow us to incorporate the parking-searching process by treating the probability that an approaching vehicle would select a given parking facility as a function of traffic and parking facility characteristics (and possibly driver characteristics). The discrete model could also potentially be more computationally efficient. Please see Appendix B for more discussions.

3.1.1. M/M/C/C queueing model

The state of the queue corresponds to the occupancy of a parking facility, defined as the number of vehicles currently parking in it. Note that parking occupancy is a discrete random variable. Denote the time-dependent probability distribution of the parking occupancy as $O(t)$, a row vector whose elements represent the probabilities of each occupancy value, from 0 to the capacity C of the parking facility. $O(t)$ is the product of initial occupancy $O(0)$ and the Markov transition-probability matrix $P(t)$.

Both arrival and departure are assumed to follow Poisson process. The arrival rate of the parking facility is denoted as λ, and the parking time of each parked vehicle follows a negative exponential distribution with a mean value of $1/\mu$ units of time. If the current occupancy is n, then the current departure rate of the entire parking lot is $n\mu$. As described above, the parking process is in fact a birth-and-death process (Ross, 2014), more specifically a Poisson queuing process with finite servers (equal to the capacity of the parking facility C) and zero buffer size. For fixed λ and μ, it is well established that the Kolmogorov's backward equations can be used to solve for the transition probability matrix $P(t)$ (Ross, 2014).

3.1.2. Model properties

This section further investigates the time-dependent model properties, such as expectation, variance, and asymptote, before applying them for parking occupancy prediction.

Let N_t denote the occupancy at time t. Assume that during a period $[t, t+h]$ where $N_t < C$ always holds; i.e., the parking facility is under-saturated. Then:

$$E(N_{t+h}|N_t) = N_t + (\lambda - N_t\mu)h$$

We have proved that (see Appendix A):

$$E_t = e^{-\mu t}\left(E_0 - \frac{\lambda}{\mu}\right) + \frac{\lambda}{\mu} \tag{1}$$

It is evident from Eq. (1) that when the facility is under-saturated, the occupancy expectation will converge to λ/μ after a sufficiently long time. Furthermore, Eq. (1) will degenerate to λt when departure rate is 0. In addition to the expected occupancy, the variance of N_t can also be written as a simple function of $\lambda,\ \mu,\ t$, if the parking facility is under-saturated (See Appendix A for detailed derivation):

$$V_t = e^{-2\mu t}(V_0 - E_0) + e^{-\mu t}\left(E_0 - \frac{\lambda}{\mu}\right) + \frac{\lambda}{\mu} \tag{2}$$

Eqs. (1) and (2) provide the basis for our parking occupancy prediction framework when the parking facility is under-saturated. First, curve-fitting techniques can be employed offline to estimate the parameters from time-series historical occupancy data. When multiple time series are available (for example, time-series occupancy data during peak hour from multiple days), the historical time-dependent mean and variance of the occupancy can be calculated to serve as the observation. Either the expectation or the variance equation would allow us to estimate both the arrival and the departure rate at the same time without additional assumptions or data requirements. Next, if current occupancy is available, expected future occupancy can be predicted from Eq. (1) with the estimated arrival and departure rates. It should be pointed out that when the facility is over-saturated, Eqs. (1) and (2) no longer hold. In this case, parameter estimation is much more difficult. We propose to adopt maximum likelihood or least squares estimation in conjecture with the Markov model directly to estimate the parameters. The two estimation approaches are discussed in the next section.

3.2. Estimating model parameters from time-series data

In this section, two methods are proposed for two different situations to estimate the parameters in the Markov models described in Section 3.1.

1. Regression parameter estimation:

   This method is applicable to the case where the capacity of a parking facility is not exceeded, i.e., the facility is under-saturated. Parameters for both the arrival and departure processes can be efficiently estimated at the same time with historical / time-series data of occupancy alone. Compared to pure statistical or machine learning methods, the governing regression equation in our approach is derived from analytical properties of the Markov models, and does not require additional tuning of the model structure (e.g., number of input nodes, number of hidden nodes, etc. in neural network approaches).

2. Direct maximum likelihood or least squares parameter estimation:

   This method can be used more generally, even when the parking facility is over-saturated. However, it is more computationally demanding. When available historical data is limited, it is recommended to fix one of the arrival and departure rates in order to obtain an accurate estimate. This is similar to the requirement in most previous studies (Caliskan et al., 2007; Caicedo et al., 2012; Portilla et al., 2009).

Numerical experiments with simulated data are first performed to demonstrate the two methods above. Case studies with real data are presented in Section 5.

3.2.1. Parameter estimation methods

3.2.1.1. Regression with derived expectation or variation equation . Suppose a parking facility is under-saturated, i.e., no arriving vehicle is rejected because the facility is full, and the arrival and departure parameters are constant during a time window $[t_s, t_e]$, where t_s is the starting time and t_e the ending time. Note that in practice, the observed occupancy is commonly in time series form. Suppose $t_e = t_s + m\Delta t$ and let $\bar{n}_{t_s+i\Delta t}$ denote the mean of the observed occupancy at time point $t = t_s + i\Delta t$, $\forall i = 0, 1, 2, \ldots m$ over multiple days. Applying Eq. (1), a constrained regression problem can be formulated to estimate (λ, μ). The objective is to minimize the square of the difference between predicted and observed mean occupancy values $Min_{\lambda,p}\ \Sigma_{i=1}^{m}(y(i) - \bar{n}_{t_s+i\Delta t})^2$, where $y(i) = e^{-\mu i\Delta t}(\bar{n}_{t_s} - \frac{\lambda}{\mu}) + \frac{\lambda}{\mu},\ \forall\ i = 1, 2, \ldots m$. This forces the fitted curve to always pass through the first data point $(t_s, \bar{n}_{t_s})$ in the time window, and eliminates error propagation when multiple time windows are being considered. The constraints assure the values of λ and μ are realistic, within the physically possible ranges of entrance and exit rates.

3.2.1.2. Maximum likelihood or least squares estimation with Markov model . When the parking facility is over-saturated, Eqs. (1) and (2) no longer hold and the regression method is no longer applicable. In this case, we propose to adopt appropriate optimization techniques directly to seek a pair of (λ, μ) for the proposed Markov model that maximizes the likelihood of observed occupancy values or minimizes the sum of error squared between observed and predicted occupancy distributions.

Suppose historical occupancy data at t_1, t_2 $(t_1 < t_2)$ from M days are available. Denoted the observed occupancy values at t_1 over M days as $n_1^1, n_1^2, \ldots n_1^k, .., n_1^M$ and those at t_2 as $n_2^1, n_2^2, \ldots n_2^k, .., n_2^M$. From these observations, the historical occupancy distributions can be calculated as $O_1 = (f_1(0), f_1(1), \ldots f_1(C))/\sum_{k=1}^{M} n_1^k$ and $O_2 = (f_2(0), f_2(1), \ldots f_2(C))/\sum_{k=1}^{M} n_2^k$, where $f_t(i)$ is the frequency of occupancy value i being observed at time t. Using O_1 as the starting point, and employing the proposed Markov model $O(t) = O(0) \bullet P(t)$ with any given pair of (λ, μ), we can calculate the predicted occupancy distribution at t_2, denoted as $\tilde{O}_2$. The maximum likelihood estimation seeks a pair of (λ, μ) that maximizes the term $\prod_{i=0}^{C} \tilde{O}_2(i)^{f_2(i)}$, the likelihood of the actual observations at time t_2 under the predicted occupancy distribution $\tilde{O}_2$. The least squares estimation, on the other hand, seeks a pair of (λ, μ) that minimizes the term $\sum_{i=0}^{C} (O_2(i) - \tilde{O}_2(i))^2$.

It is worth noting the time-dependent occupancy distribution from the continuous-time model is difficult to evaluate in general. It involves evaluation of a matrix exponential operator $e^X := \sum_{k=0}^{\infty} X^k/k!$ (Coddington and Levinson, 1955), which is non-trivial and does not have a closed-form expression when the arrival or departure rate varies with time. Other reasonable queuing models can also be adopted in the framework. Appendix B discusses one of such alternatives, a discrete-time model, which could potentially be more computationally efficient.

3.2.2. Numerical examples

To demonstrate the occupancy prediction framework, simulation experiments are performed. Simulated arrivals and departures are generated based on the parking model, and the resulting occupancy values are treated as the observed historical data. Note that since the process used to generate the random data is identical to the one used in the estimation approaches, the numerical examples in this subsection do not necessarily validate the proposed models for real-world application. Rather, the simulated numerical examples explore the applicability of the two parameter estimation methods in under- and over-saturated conditions. The appropriateness of the proposed methods in real-world applications is demonstrated in Section 5.

In the simulation, the capacity of the parking facility C is set to 20. The simulation duration is 50 time intervals. The average arrival rate λ is 1 vehicle per time interval, and the departure rate μ is set to 0.05 (equivalent to an average parking time of 20 time intervals). The starting occupancy distribution has non-zero probabilities for occupancy values 2

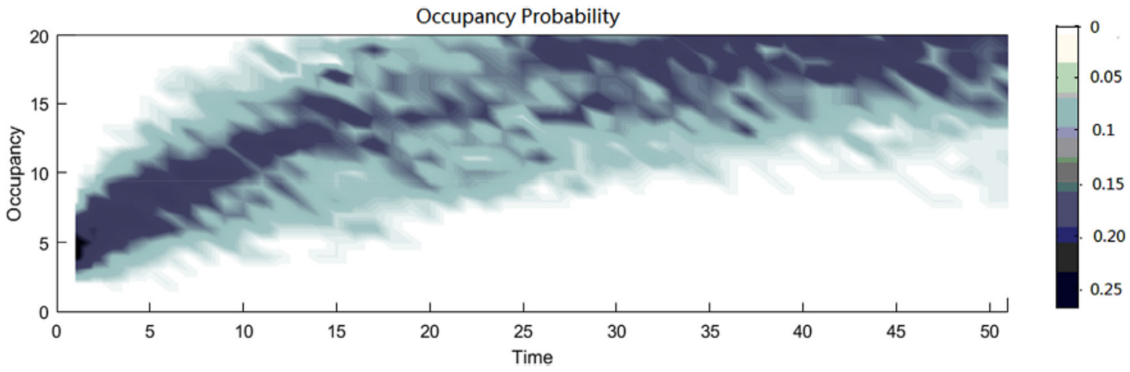

Fig. 1. Occupancy distribution over time from simulated data.

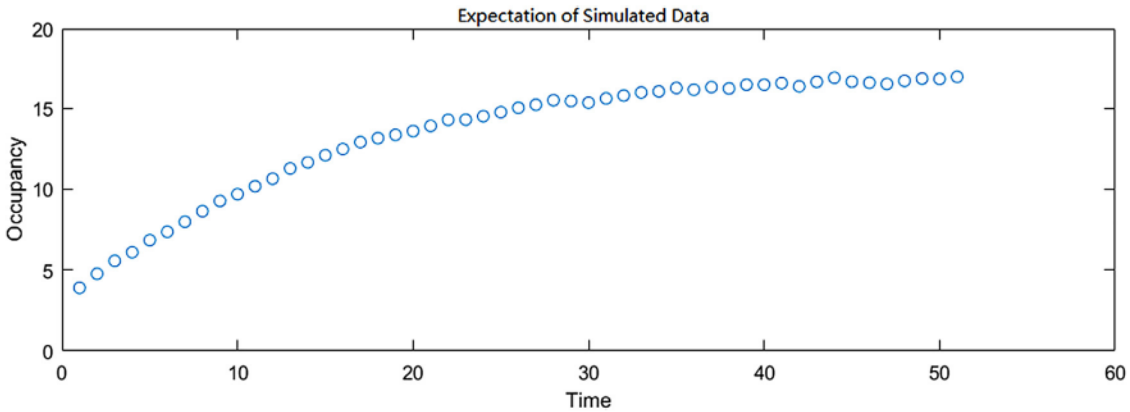

Fig. 2. Occupancy expectation over time from simulated data.

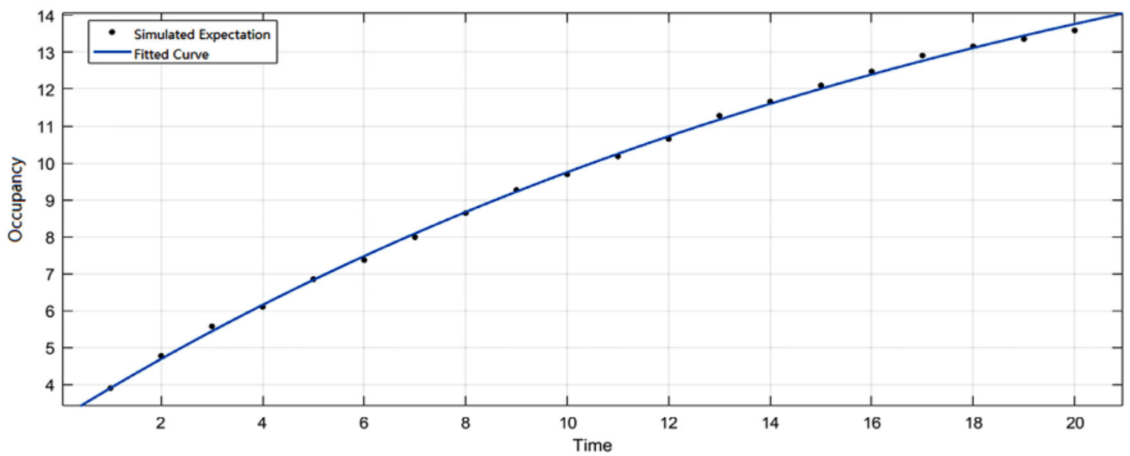

Fig. 3. Curve fitting for $t \in [1,\ 20]$.

– 6 as 0.1, 0.3, 0.3, 0.2, and 0.1, respectively. From simulated arrivals and departures over 100 simulation runs, the distribution of occupancy (Fig. 1) and the expectation of occupancy (Fig. 2) over time are computed to serve as historical data.

3.2.2.1. Under-saturated facility. From Fig. 1, it can be seen that the occupancy rarely reaches capacity when $t \leq 20$. Therefore, we can apply the regression method for the period [1, 20]. Fig. 3 shows the fitted expectation curve. The fitted parameters $(\lambda,\ \mu)=(1.0175,\ 0.0521)$ from 20 data points are very close to the simulation input. It is worth mentioning that gate constraints are not explicitly enforced in this simulated experiment.

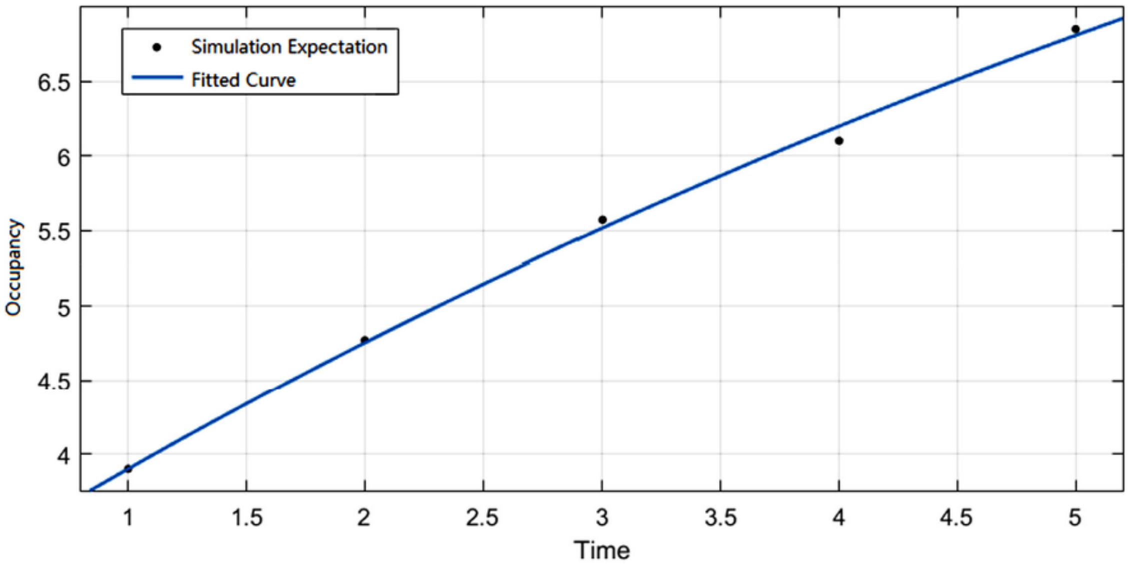

Fig. 4. Curve fitting for $t \in [1,\ 5]$.

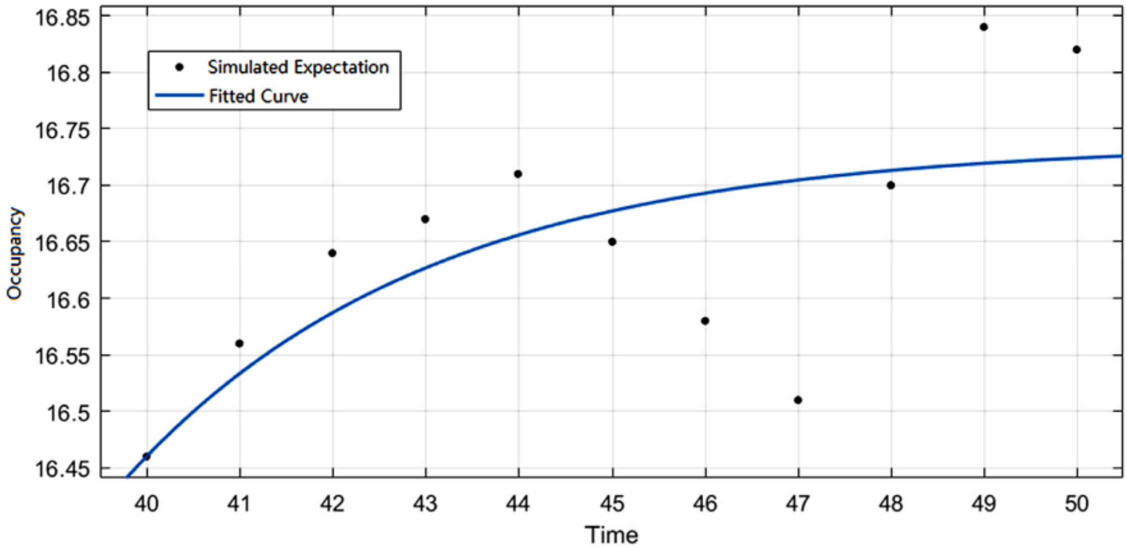

Fig. 5. Curving fitting for $t \in [40,\ 50]$.

In real application, arrival and departure parameters may not remain constant for an extended period, which means that there might only be a limited number of data points for curve fitting. To see how the regression method performs when there are fewer data points, an expectation curve is fitted with the first 5 data points only (Fig. 4). The estimation result is $(\lambda,\ \mu)=(1.0322,\ 0.0533)$, still quite close to the simulation inputs. This indicates that regression method could work well even with fewer data points.

We also performed estimation using the variation curve Eq. (2) and obtained very similar results.

3.2.2.2. Over-saturated facility. In this simulation experiment, there is a high probability that the parking facility is fully occupied as t increases (Fig. 1). In this case, the expectation curve Eq. (1) no longer holds. If the regression method is applied nonetheless, the results would be inaccurate. The estimated $(\lambda,\ \mu)=(2.7833,\ 0.1645)$ using data from $t \in [40,\ 50]$ (Fig. 5) is quite different from the simulation input.

Instead, direct maximum likelihood and least squares estimation is implemented with a compass search algorithm (Kolda et al., 2003) to seek the best feasible pair of $(\lambda,\ \mu)$. Both λ and μ can be estimated simultaneously rather accurately, if the observation data set is large (for example, with 10,000 time series). However, when the size of historical data is limited (such as in our case with 100 time series), the direct optimization may converge to a local optimal depending on

Table 1
Simulation-based optimization estimation results for $t \in [40, 50]$.

Fixed departure rate $\mu = 0.05$ Estimate λ			Fix arrival rate $\lambda = 1$ Estimate μ		
Starting point	Maximum likelihood	Least square	Starting point	Maximum likelihood	Least square
0	0.9974	0.9973	0	0.0486	0.0494
0.1	0.9973	0.9973	0.2	0.0493	0.0500
0.2	0.9973	0.9973	0.1	0.0488	0.0488
0.5	0.9974	0.9973	0.067	0.0500	0.0500
1	0.9974	0.9973	0.05	0.0500	0.0500

the starting point of the compass search. In this case, it is ideal if one parameter can be fixed based on prior knowledge. For example, with gated parking facilities, the arrival rate can be obtained from gate readings. Table 1 reports the estimation results from simulation-based optimization with various starting values and different objectives, when one of the two parameters is fixed. It can be seen the estimated values are quite accurate, regardless of the starting point.

3.3. Parameter estimation in real-world applications

Before applying the regression method, we need to identify the days from all available data when drivers have similar activity patterns that will result in similar arrival rate and parking time. With those groups defined, a parking facility is considered to experience the same arrival and departure rate at the same time of day for any day in a given group. On the other hand, it is important to determine the time of day when λ and μ change. Clustering techniques can be employed for the former; and an iterative approach can be adopted for the latter.

3.3.1. Day-to-day patterns

To avoid over complicating and dividing historical data into too many groups, we propose to only consider two to three groups of days, using between-group linkage (Murtagh, 1983) hierarchical cluster method in SPSS (IBM, 2016). Possible classifications include “workdays” versus “holidays”, or “regular workdays” versus “high demand days” versus “holidays”.

Between-group linkage hierarchical cluster is a common clustering method. Each sample is initially treated as a separate group. The gaps between every pair of groups are calculated and the two closest groups are combined into one. The gaps between groups are defined as the average distance between each pair of data points in different groups; and the distance between data points is defined as the Euclid distance. This process is repeated until the desired number of groups is achieved.

3.3.2. Within-day patterns

We define the time of day when λ and μ change as breaking points. Apparently, $(\lambda, \ \mu)$ must change at local maximum or local minimum of the historical mean occupancy curve. These breaking points divide the total analysis period into multiple periods within which the historical expectation curve is monotone. For a period where the historical mean occupancy decreases, it is obvious that $\mu > 0$; and the least square estimation from 3.2.1 should be applied for parameter estimation. When historical mean occupancy monotonically increases (μ may or may not be zero), both the original exponential and degenerated linear formulations can be used to fit the data and the one with better performance chosen. Furthermore, since parameters normally do not remain constant for an extended period in real world, we let m represent the largest number of time intervals that $(\lambda, \ \mu)$ could remain constant, and start our regression with a time window of at most m data points. If the resulting R^2 is sufficiently large (greater than a pre-defined value), move to the next period. Otherwise, reduce the length of the period by one time interval at a time until R^2 meets the criteria. A flowchart of the process is shown in Fig. 6.

3.4. Occupancy prediction in real-world applications

Once the time-dependent arrival and departure rate and are estimated, they can be applied to predict future occupancy using Eq. (1). The prediction module can be engaged both with and without real-time updates, and does not require additional data other than observed occupancy at the beginning of the day or in real time. Additionally, it would be beneficial to monitor the prediction error in real-time, although not required, in order to make any adjustments to capture abnormal parking demand caused by special events or other unforeseen reasons that may not be reflected otherwise in the historical data set.

3.4.1. Prediction with and without real-time updates

The model-based prediction method proposed allows occupancy prediction for any time in the future both with and without real-time data.

Without any real-time occupancy data, prediction of the next time interval can be performed repeatedly based on the output from previous occupancy prediction. For example, a 24-h prediction without real-time data predicts the occupancy

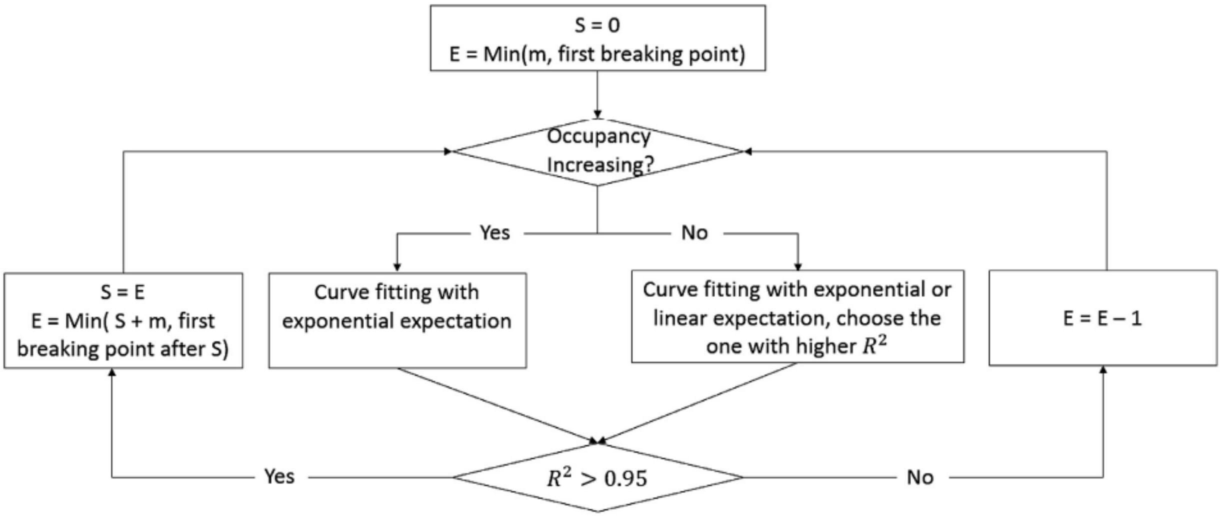

Fig. 6. Iterative estimation framework.

for the entire day based on the observed occupancy at the beginning of the day. It is certainly limited in handling abnormal arrival and departure patterns, but has its use in planning and other transportation analysis.

We can also update the prediction based on real-time data as they become available. The prediction interval therefore could be significantly shorter, bounded only by the data collection interval. Note that this does not mean that the only prediction available to users is from the previous time interval. Rather, predicted parking occupancy is available whenever a user queries the information, be it at the beginning of the trip or anytime during the trip. The real-time updates allow travelers to check for changes in occupancy prediction shortly before they arrive at their destination for a more accurate estimate.

3.4.2. Recurrent over-saturation and special event

As described in Section 3.2.1, if a parking facility is recurrently over-saturated, a direct maximum likelihood or least squares estimation can be adopted. However, if a parking facility is only occasionally oversaturated, possibly due to special event, then the direct estimation methods would not be applicable. This is because the near- or over-saturation is not caused by recurring arrival and departure patterns, and there would not be sufficient historical data to support the direct estimation. There are multiple possible approaches to handle special event occupancy prediction, depending on data availability.

If rich historical special event data (days, times, durations, and types, etc.) is available, it is possible to apply clustering methods to divide the historical data into groups and learn the arrival rate and departure probability for each group. It is worth noting that care should be taken in accounting for the normal background parking demand during special events. For prediction, if information regarding the special event of question is known in advance, then one or more groups that have similar attributes could be selected to carry out feature-weighted average method (Tamrazian et al., 2015) and alike. Otherwise, multiple prediction threads based on different groups could be performed simultaneously, and their prediction errors monitored in real-time. The weight of each group can then be updated over time as a function of the prediction errors.

If information on historical special event is insufficient, i.e., the underlying parking searching and queueing processes are not well-understood, reasonable predictions can still be achieved by refining the estimates based on real-time prediction error. When the actual occupancy becomes greater than the predicted value by a pre-defined threshold, it indicates that the historical arrival and departure rates are not able to capture the parking demand pattern of the moment, and that the ingress of a special event may be in process. The real-time prediction errors could serve as a simple piece-wise linear approximation to the additional special event arrivals. The occupancy prediction can be slightly modified by adding this component to the value predicted from the historical arrival and departure patterns. Note that in this approach, we do not need to know whether a special event is present or not, nor its time and duration.

4. Case studies

Three case studies (two for regular workdays and one for a special event) with real occupancy data from San Francisco are presented to validate the performance of the proposed framework. These case studies were originally done with discrete-time model but as described above, the occupancy prediction will not be different for continuous-time model.

It is worth mentioning that for all the 14 parking facilities we collected data for, none was ever filled during our three-month data collection period. However, there were days where some parking facilities were very close to saturation due

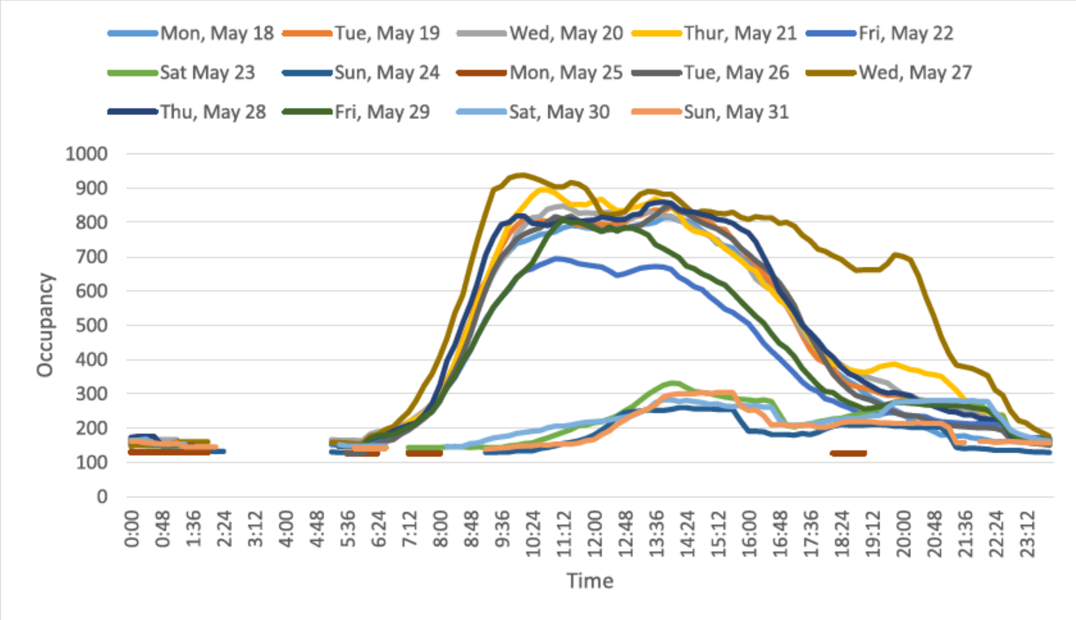

Fig. 7. Occupancy data (Civic Center Garage, San Francisco, May 18–31, 2015).

to special event. In view of this, all three case studies in this section have adopted the regression method for parameter estimation. The prediction (both with and without real-time update) for regular workdays is straightforward. The case studies show that the proposed model-based framework performs well both with and without real-time updates for a range of prediction time windows, where the paradigm shift from daytime to evening can be captured by the prediction with real-time updates. The special event case study demonstrates that the framework can achieve accurate prediction even if the underlying model may be inaccurate due to limited data availability.

4.1. Data collection

A Hypertext Preprocessor (PHP) script was written to retrieve the San Francisco parking garage JavaScript Object Notation (JSON) file from Firebase's (Firebase, 2015) public data set from May 2015 to August 2015. A Cron job has been established to execute the PHP script every two minutes to keep bandwidth use low on the Firebase servers. The script parsed the garage name, the number of open spaces and the number of total spaces from the JSON file before writing the parsed data and a timestamp to a Comma Separated Values (CSV) text file. The San Francisco garage parking data was retrieved with the JSON file. No street parking data was available during our data collection period.

4.2. Case study 1

The purpose of this case study is to demonstrate all the components of the proposed framework. The case study is based on two weeks' worth of data from the Civic Center garage collected in May 2015. Fig. 7 plots the observed occupancy every 12 min for the two weeks. The missing points in the figures are due to closure of the garage and lost connection to the database. These points are excluded from the occupancy expectation calculation. Due to the low occupancy between midnight and 6 a.m., we will focus on parameter estimation and occupancy prediction starting from 6 a.m.. In view of the lack of any additional information on the maximum entering and leaving rates, we set both g_1 and g_2 to 60, i.e., it is assumed that at most 60 vehicles on average could enter or leave the parking garage every 12 min.

From hierarchical cluster analysis (Fig. 8), the 14 days are first divided into two groups: "workday" and "holiday" . Notice that May 25, 2015 is a weekday but is classified into the "holiday" group, because it was Memorial Day. Also, May 27 displays a significantly different occupancy pattern from other workdays, possibly caused by special events or randomness. Therefore, we excluded May 27, 2015 and used the rest of the days in the "workday" group as our analysis data set (a total of 8 days).

The expectation curve of the model (Eq. (1)) is employed to estimate and predict "workday" group occupancy data for the Civic Center garage. The longest time window during which (λ, μ) remain constant is set to 1 h in this case study. If R^2 of the regression is smaller than 0.95, the time window will be reduced by 12 min until R^2 is sufficiently large.

Table 2 reports the estimated parameters during 13:00 and 15:36. From 13:00 to 14:00, the mean occupancy from the 8 days monotonically increases. Both exponential (shown as Exponential in Table 2) and linear (shown as Linear in Table 2)

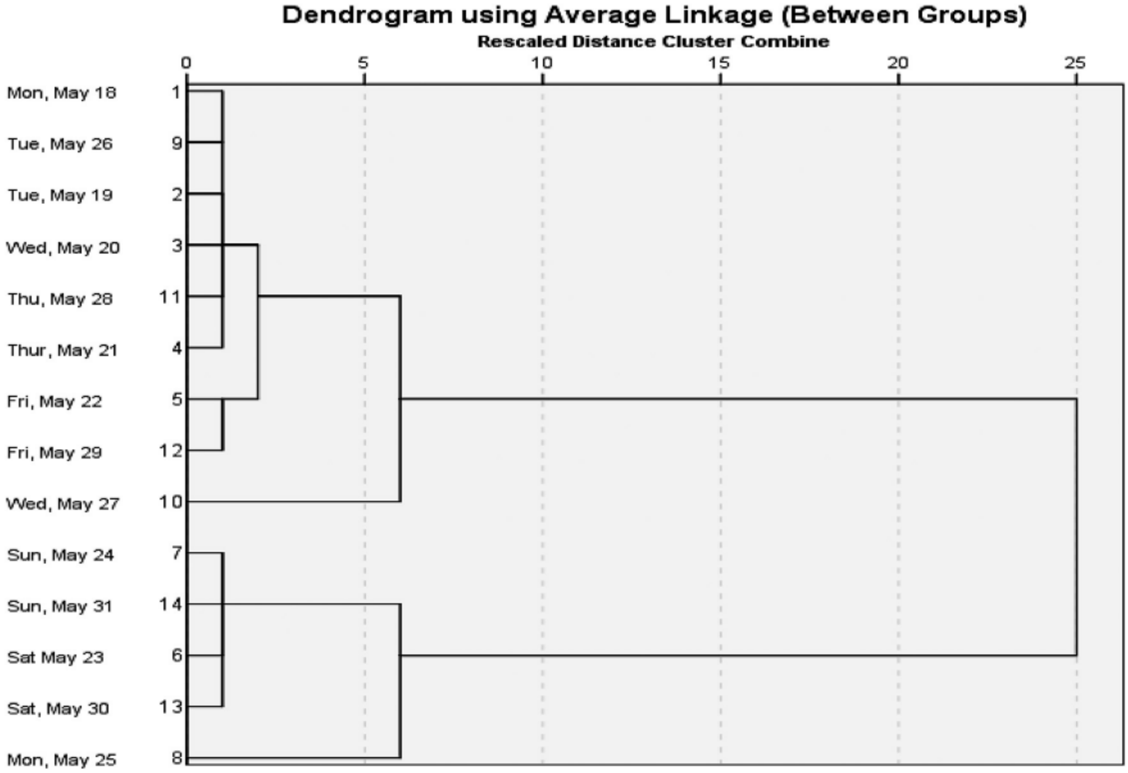

Fig. 8. Hierarchical cluster.

Table 2
Part of estimated parameter table.

Interval	Starting time	Historical mean occupancy	Exponential		Linear
			λ (# vehicles every 12 min)	μ (leaving rate for every parked vehicle)	λ (# vehicles every 12 min)
1	13: 00	811.5000	–	–	6.8
2	13: 12	817.3333	–	–	6.8
3	13: 24	828.8333	–	–	6.8
4	13: 36	835.0000	–	–	3
5	13: 48	838.0000	–	–	0.167
6	14: 00	838.1667	0	0.0133	Na
7	14: 12	823.1667	0	0.0133	Na
8	14: 24	814.0000	0	0.0133	Na
9	14: 36	802.8333	0	0.0133	Na
10	14: 48	792.0000	0	0.0133	Na
11	15: 00	781.8333	0	0.0233	Na
12	15: 12	767.5000	0	0.0233	Na
13	15: 24	755.0000	0	0.0233	Na
...	...	...	...	...	...
22	17:12	533.6667	54.2751	0.1883	Na
23	17:24	488.5000	54.2751	0.1883	Na
24	17:36	457.6667	54.2751	0.1883	Na
25	17:48	428.6667	54.2751	0.1883	Na
26	18:00	404.8333	54.2751	0.1883	Na
27	18:12	382.8333	80.9103	0.2723	Na
28	18:24	362.0000	80.9103	0.2723	Na
29	18:36	346.0000	70.82864	0.2384	Na
30	18:48	336.1667	70.82864	0.2384	Na
31	19:00	326.1667	70.82864	0.2384	Na
32	19:12	318.5000	0	0.0219	Na
33	19:24	311.8333	0	0.0219	Na
34	19:36	305.1667	0	0.0219	Na

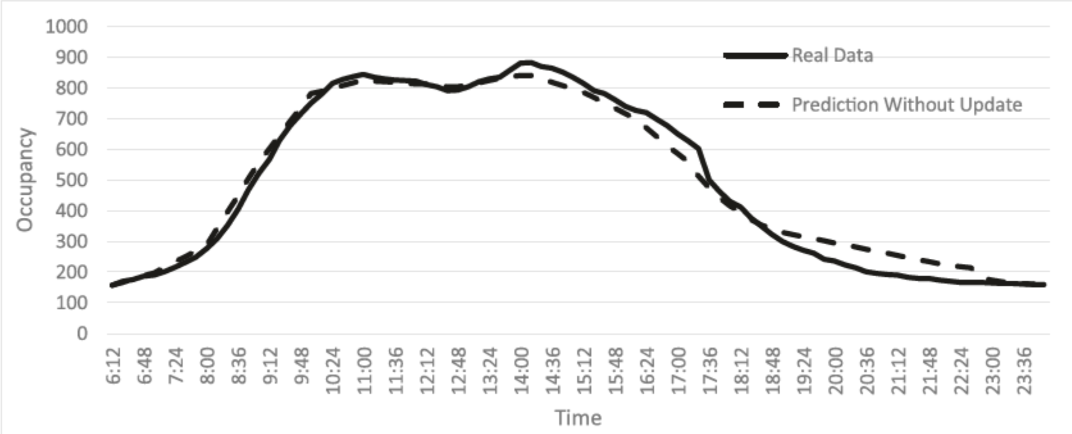

Fig. 9. Prediction without update (Civic Center Garage, San Francisco, June 2, 2015).

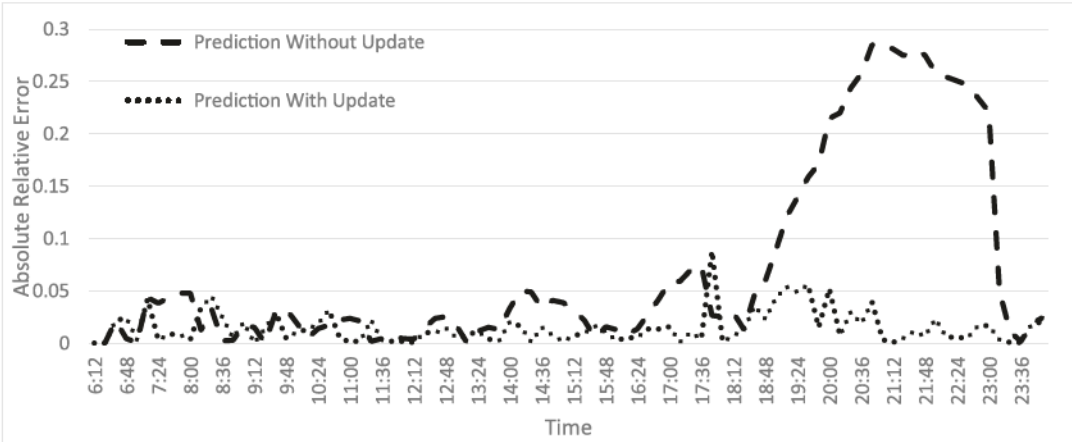

Fig. 10. Absolute relative prediction error with and without update (Civic Center Garage, San Francisco, June 2, 2015).

formulations are employed; and it turns out that the Linear model, where the leaving probability is set to zero, fits the data better. Also note that the data did not support the assumption that the arrival rate is constant between 13:00 and 14:00. Rather, the final estimation results suggest that the arrival rate remained constant from 13:00 to 13:36, and kept decreasing during 13:36 to 14:00. Starting from 14:00, the mean historical occupancy began to decrease, and the Exponential model should be applied. The results show that 15:00 is a breaking point where the leaving probability almost doubled.

With the estimated time-dependent arrival rate and leaving probability for workdays from the last two weeks of May 2015, we performed prediction every 12 min with and without real-time occupancy data (updated every 12 min). This case study uses testing data from June 2, 2015 to validate and demonstrate the performance of the proposed framework and methods.

The prediction without update is based on the observed occupancy at 6:00 a.m.. The results are presented in Figs. 9 and 10. The absolute relative error of the prediction is less than 10% from 6 a.m. to 7 p.m. From 7 p.m. to 10 p.m., however, the occupancy is over-estimated (Fig. 9), and the relative error is significant (Fig. 10). A closer look at the event calendars of venues at Civic Center revealed that June 2nd, 2015 had no event after 5 p.m. while most of the days in the historical data set had at least one event after 7 p.m.[1] This explains the over-estimation in the evening period for June 2nd, 2015, and also highlights the limitation of prediction without update.With real-time data, an updated prediction is produced every 12 min in this case study. As can be seen from Fig. 10, prediction with update is able to capture the paradigm shift from daytime to evening, and produce estimates that are more accurate with a maximum absolute relative prediction error of less than 9% and a mean absolute relative error (MARE) of 1.486% for the entire day.

[1] http://www.sfstation.com/calendar/san-francisco/civic-center/06-02-2015.

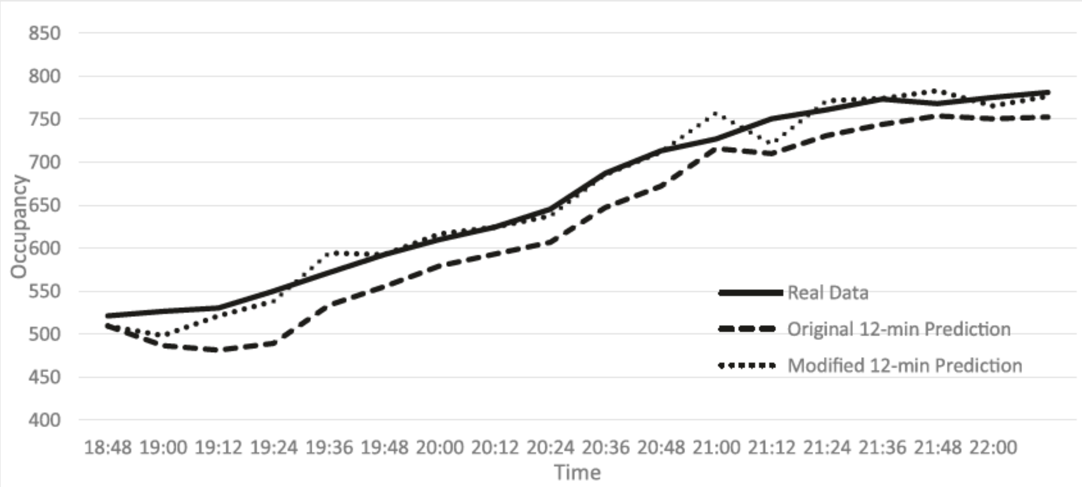

Fig. 11. Modified vs. original 12-min prediction (Civic Center Garage, San Francisco, Jun 19, 2015).

4.3. Case study 2

The purpose of this case study is to examine the performance of the proposed model-based prediction framework when a parking facility is highly congested. Although none of the 14 parking garages was ever full during our data collection period, the data set does have days where a parking facility was close to saturation due to special event. June 19th, 2015 stands out for the Civic Center parking garage with abnormally high parking occupancy. It reached an occupancy of over 90% in the evening. The peak occupancy is 781 out of a capacity of 843 at 10:12 p.m. This indicates that the facility could be nearly- or even over-saturated on that day. Although the data shows that there were still some stalls available, they were not necessarily usable due to various reasons (reserved stalls, double-parking, undesirable surroundings such as low beams or support columns, etc.) Looking back, it is evident that the congestion was caused by the 100th celebration of San Francisco's City Hall on that day.[2] Since data on historical events of this scale is insufficient to perform the direct estimation method described in Section 3.2 or the clustering method described in Section 3.4.2, the modified prediction method as described in Section 3.4.2 is adopted for this case study.

We focus on a portion of the ingress window, which is from 6:48 p.m. to 10:12 p.m., where the occupancy kept increasing. Since June 19th, 2015 is a Friday, data from the following three Fridays was used as our training dataset. We obtained the estimated arrival rate and departure probability for a "normal" Friday using the same estimation procedure as in the previous case study. The parameter estimation time interval Δt is 12 min, and the maximum time window during which (λ, μ) remain constant is also set to 12 min. Both the original and the modified prediction are performed, where the threshold is set to 10.

Fig. 11 shows the predicted occupancy for the two prediction methods. Without any training data from comparable special events, the linear approximation in the modified prediction method is able to handle the extra arrival well. Comparing to the original prediction, the modified method reduces the MARE during the study window from 5.23% to 1.74%, which is comparable to the MARE from the normal workday scenario in Case Study 1.

4.4. Case study 3

The purpose of this case study is to investigate how different prediction intervals affect the prediction accuracy. The case study is based on historical data from July 6 to July 15, 2015. The procedure is the same as in case study 1. Since data was collected every 2 min during this period, the time interval Δt for parameter estimation is therefore set to 2 min. The maximum time window during which (λ, μ) remain constant is set to 40 min.

Occupancy prediction for 2, 4, 6, 8, 10 and 12 min in the future is performed based on the current real-time data. As expected, prediction error increases as the prediction interval gets longer. Fig. 12 shows the errors of the predictions made 2, 6, and 12 min earlier. The evening period is selected because the errors are insignificant during the day even for the longest prediction interval. For predictions made 12 min ago, the absolute relative error is no more than 8% during the evening period. The value reduces to 5% and 4% for predictions made 6 and 2 min ago, respectively. This suggests that while a short prediction interval leads to more accurate estimates, predictions made a while ago is still effective.

---

[2] https://www.sfstation.com/calendar/san-francisco/06-19-2015.

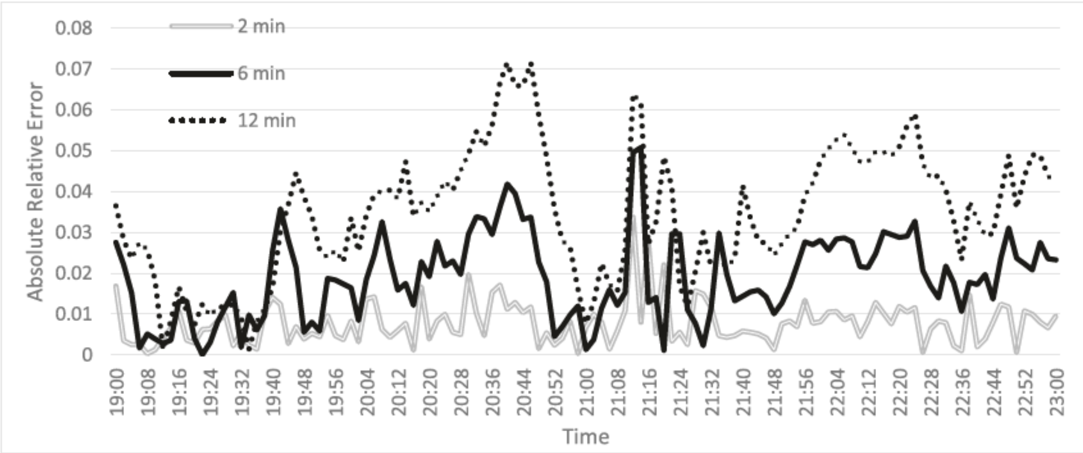

Fig. 12. Absolute relative prediction error for three different prediction intervals (Civic Center Garage, San Francisco, July 16, 2015).

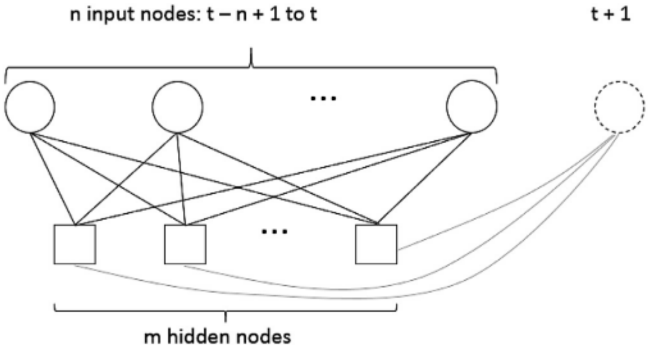

Fig. 13. Neural network structure for prediction of interval $t+1$.

5. Comparison with machine learning methods

To gain further insight into the applicability and suitability of the proposed model-based framework, this section compares the performance of our framework to those from three pure machine-learning methods found in recent literature. They include artificial neural network (ANN, Vlahogianni et al., 2016), wavelet neural network (WNN, Ji et al., 2015), and feature-weighted average method (Tamrazian et al., 2015). The former two are applicable to prediction with real-time update and the last is for prediction without real-time update only. These methods are selected primarily because they require the same data input as in our framework.

Vlahogianni et al. (2016) and Ji et al. (2015) both proposed to predict the occupancy for time interval $t+1$ using real-time data from n past time intervals (including the current interval t). The general structure of their models is shown in Fig. 13. The ANN (Vlahogianni et al., 2016) employs a sigmoid function at each hidden nodes to transform the weighted sum of all input nodes, whereas the WNN (Ji et al., 2015) uses Morlet function instead.

The number of input nodes n, number of hidden nodes m, and the learning rate and momentum have to be determined before the actual learning of weights of each input and hidden node can take place. Vlahogianni et al. (2016) optimized the ANN structure using genetic algorithm with 5-fold cross-validation (Tan et al., 2005). This process involves partitioning the original data randomly into five subsets of the same size. Given a neural network structure, it is evaluated five times, using each of the five subsets as testing data and the rest as training data for the weights of each input and hidden node. The average performance over the five evaluations of the neural network structure is used as the fitness measure in the genetic algorithm. Ji et al. (2015) did not provide details on how the parameters relevant to model structure are determined other than that they were selected by experiments.

The feature-weighted average method (Tamrazian et al., 2015) clusters the time series of daily occupancy into different groups. The prediction is performed by a weighted average of the means of each group, where the weights are proportional to the number of days with the same feature as the target day (day of week, weather etc.) in different groups, and does not require real-time data.

Comparisons based on the case studies described in Section 4 show that our approach significantly outperforms the wavelet neural network and feature-weighted average methods in terms of prediction error. Artificial neural network is able

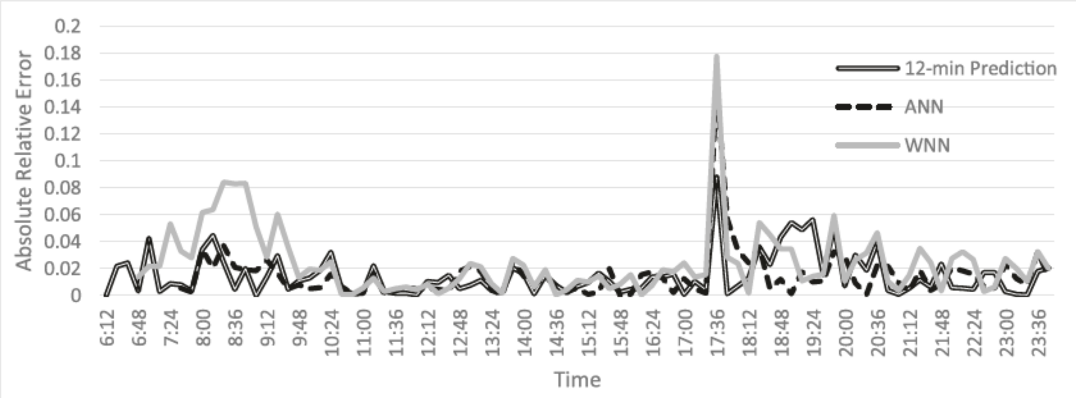

Fig. 14. Relative error of three methods for case study 1.

Table 3
MARE for prediction with update for the next interval.

	Approach	Case 1	Case 2
Proposed in this paper	(w/ Original prediction with update)	1.486%	5.23%
	(w/ Modified prediction with update)	–	1.74%
	ANN	1.464%	2.81%
	WNN	2.413%	–

to achieve the same level of accuracy comparing to our approach but requires hours to tune the model structure before training, whereas our approach only takes less than 30 s to estimate the arrival and departure parameters.

5.1. Prediction with real-time update for the next interval

This subsection first investigates the performances of ANN, WNN, and our model-based approach in predicting the occupancy for the next time interval, as proposed in Vlahogianni et al. (2016) and Ji et al. (2015). We trained separate ANN and WNN models for both case studies 1 and 2. For the ANN models, we implemented the same 5-fold cross validation genetic algorithm used in Vlahogianni et al. (2016) to optimize the ANN structure, including the number of input nodes (number of previous intervals), the number of hidden nodes, learning rate and momentum. Following the same setting as in Vlahogianni et al. (2016), 50 chromosomes and 100 generations are used in the genetic algorithm. Our implementation is based on Java with Weka API (Hall et al., 2009). On the other hand, since it is not clear how exactly the WNN model was tuned in Ji et al. (2015) and no well-established package for WNN exists, we programmed a brute-force procedure in Matlab 2017a to determine the WNN structure. Once the neural network structure is determined, the weights for each input and hidden nodes are then trained using the full historical data sets used in Sections 4.2 and 4.3.

Fig. 14 shows the absolute relative error from ANN, WNN, and our approach for Case Study 1. It can be observed that our method performs considerably better than WNN (especially during the morning peak) and is comparable to ANN. It is worth noting that the errors from all three methods spiked when the daytime to evening transition occurred. For the overall performance, WNN leads to a MARE of 2.413%; ANN enjoys the lowest MARE of 1.464%; and our method is at 1.486% (see Table 3). It is not clear whether the differences are statistically significant or not for this case study, since the number of data points in these time series is relatively small. However, while our approach took less than 30 s to estimate the arrival rates and departure probabilities, it took 8.5 h to tune the structure of the ANN.

Fig. 15 plots the absolute relative prediction error from our approach (with both original and modified prediction) and ANN for Case Study 2. The same model tuning and training processes are employed. In this case study of a high profile special event, the training data is limited as data from comparable special events does not exist. Because of this, both our original prediction and ANN suffered from relatively large MARE compared to that in Case Study 1, where the time interval is also 12 min. However, with modified prediction, our approach is able to predict the occupancy more accurately with a MARE of 1.74%, much lower than ANN and the original prediction (See Table 3).

5.2. Prediction with real-time updated for multiple intervals in the future

This section examines how our method compares with ANN when predicting for a future time point that is multiple intervals ahead. This is of practical importance because travelers are likely to query ahead of time in addition to obtaining updates right before arrival. In our approach, prediction can be performed for any time point in the future using the same model with the time-dependent arrival and departure rates already estimated. For ANN, there are two ways to predict the

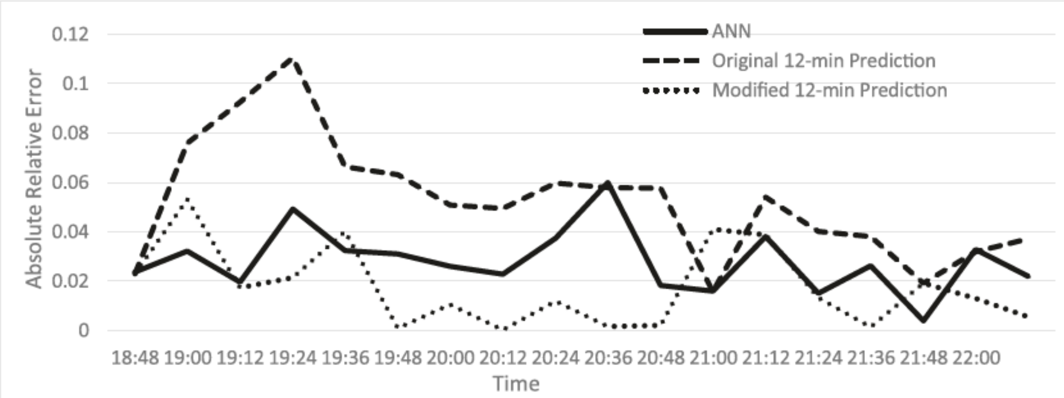

Fig. 15. Relative error of three methods for case study 2.

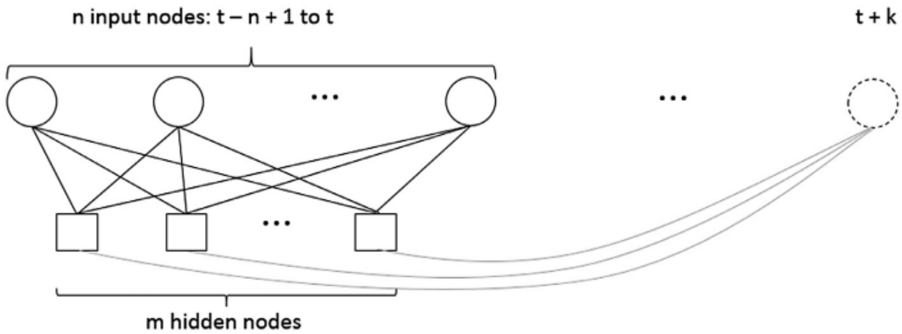

Fig. 16. Neural network structure for prediction of interval $t+k$.

Table 4
MARE for different prediction intervals.

Prediction interval (min)	2	4	6[a]	8[a]	10[a]	12[a]
Our approach	0.489%	0.812%	1.061%	1.283%	1.482%	1.672%
Repeated application of the next-interval ANN	0.486%	0.867%	1.303%	1.803%	2.378%	2.998%
Direct ANN for 12 min in the future	–	–	–	–	–	1.697%

[a] The MAREs from repeated application of the next-interval model and from our approach are statistically different at 99% confidence level.

occupancy for time $t+k$ at the current time t: 1) applying the next-interval model for $t+1$ (as shown in Fig. 13) repeatedly for k times, using predicted values as part of the inputs; and 2) applying a direct model for $t+k$ (see Fig. 16), which needs to be trained separately beforehand. It is expected that a direct ANN model for $t+k$ would outperform repeated application of the next-interval ANN model. However, since k is not fixed (travelers could query for any time point in the future), training a separate model for each possible k value could be time-consuming.

Using data from Case Study 3, where the real-time occupancy is collected every 2 min, we compare the performance of our method and ANN (repeated application of a next-interval model) for various prediction intervals ranging from 2 to 12 min. A direct ANN model to predict for 12 min in the future is developed separately and compared against our approach as well.

The resulting MAREs are shown in Table 4. When repeatedly applying the next-interval model, the MARE from ANN increases faster than our approach as the time point of interest gets further out in the future. While the next-interval ANN model achieves statistically similar performance compared to our method for 2 and 4 min in the future, its MAREs become statistically larger than those from our approach starting from the prediction for 6 min in the future. For the prediction for12 min in the future, the MARE from repeatedly applying the next-interval ANN is almost twice that of our approach. On the other hand, the direct ANN model specifically trained to predict for 12 min in the future is able to achieve the same performance as our approach.

In terms of computation time, since Case Study 3 has significantly more data than Case Study 1, it is expected that it would take days to tune the ANN using the same genetic algorithm setting in Vlahogianni et al. (2016) (50 chromosomes and 100 generations). Instead, we performed the genetic algorithm with 10 chromosomes and 10 generations, and the cor-

Table 5
MARE for prediction without update methods.

Approach	Case 1	Case 3
Proposed in this paper	8.907%	6.363%
Feature-weighted average	10.676%	7.680%

responding tuning time is 1 h 37 min for the next-interval model and about the same for the direct model for 12 min in the future. Parameter estimation in our approach took less than 2 min.

5.3. Prediction without real-time update

For prediction without real-time data, we compared our method with the feature-weighted average method in Tamrazian et al. (2015). The resulting MAREs for both case studies 1 and 3 are shown in Table 5. Our approach leads to considerably lower MAREs in both cases. A two-sample T-test also confirms that the differences are statistically significant.

6. Discussion and conclusion

This paper proposes a model-based practical framework to predict future parking occupancy from historical occupancy data alone. Being model-based rather than purely statistical, the parameter estimation methods and parking occupancy prediction in the predictive framework benefit from time-dependent analytical properties of the underlying queuing model and are computationally efficient. In addition, several practical considerations for real world implementation are accounted for in the framework, with methods proposed to handle variations of arrival and departure patterns from day to day and within a day, including special events. Using both simulated and real data, we have demonstrated that the proposed framework and methods are able to effectively estimate the arrival and departure rates offline from historical occupancy data alone, and accurately forecast parking availability with and without real-time occupancy data. It is found that our approach delivers equal or better performance compared to several pure machine-learning methods from recent literature, but requires the computation time that is orders of magnitude less to tune and train the model. Additionally, our approach can predict for any time in the future with one training process, while machine-learning methods have to train a specific model for a different prediction interval to achieve the same level of accuracy.

While this study explored methods for parking occupancy prediction when a facility is congested due to abnormal parking demand (such as special event), the dataset we currently have do not seem to be from facilities that are recurrently oversaturated. We plan to collect more data from additional facilities such as on-street parking facilities that are more likely to be recurrently over-saturated. If we are successful in obtaining such data, our future research will investigate how appropriate the proposed direct estimation method (as described in Section 3.2.1) is for such facilities. Additionally, the model-based methods in our framework allow us to tie in the parking availability estimation with network traffic and driver behaviors through the arrival and departure rates. This study is the first step towards being able to incorporate the parking searching process (reflected in the stochastic arrival process) in parking occupancy prediction, as well as to investigate network traffic effects as a result of parking availability information provision and the subsequent possibly altered parking search behaviors using game-theoretical approaches.

Acknowledgments

This research is supported by National Science Foundation project CMMI 1363244 “Collaborative Research: Modeling and Analysis of Advanced Parking Management for Congestion Mitigation”. We would like to thank the anonymous reviewers for their constructive comments and suggestions that helped improve this paper substantially.

Appendix A. Expectation and variance of the occupancy in the continuous-time model

Let h denote an infinitesimal amount of time, N_t the current occupancy. The following transition probability equations hold (Ross, 2014):

$$P(N_{t+h} = N_t + 1) = \lambda h$$

$$P(N_{t+h} = N_t - 1) = N_t \mu h$$

$$P(N_{t+h} = N_t) = 1 - (\lambda + N_t \mu)h$$

$$P(N_{t+h} = i) = 0, \ \forall i \in N, \ i \neq N_t, N_t - 1, \ N_t + 1$$

From the equations above, we have:

$$E(N_{t+h}|N_t) = N_t + (\lambda - N_t\mu)h$$
$$\Rightarrow E_{t+h} := E(N_{t+h}) = E_t + (\lambda - E_t\mu)h$$

When $h \to 0$, the above equation becomes:

$$E_t' = \lambda - \mu E_t$$

The solution to the partial differential equation above is:

$$E_t = e^{-\mu t}\left(E_0 - \frac{\lambda}{\mu}\right) + \frac{\lambda}{\mu}$$

Similar to the derivation in Appendix B, the variance can be calculated from conditional expectation and variance.

$$V(E(N_{t+h}|N_t)) = V(N_t + \lambda h - \mu h N_t) = (1-\mu h)^2 V_t$$
$$E(V(N_{t+h}|N_t)) = E(\lambda h + \mu h N_t) = \lambda h + \mu h\left(e^{-\mu t}\left(E_0 - \tfrac{\lambda}{\mu}\right) + \tfrac{\lambda}{\mu}\right)$$
$$\Rightarrow V_{t+h} := V(N_{t+h}) = E(V(N_{t+h}|N_t)) + V(E(N_{t+h}|N_t)) = (1-\mu h)^2 V_t + \lambda h + \mu h\left(e^{-\mu t}\left(E_0 - \tfrac{\lambda}{\mu}\right) + \tfrac{\lambda}{\mu}\right)$$
$$V'_t = \lim_{h\to 0} \tfrac{V_{t+h} - V_t}{h} = \lim_{h\to 0} \tfrac{(1-uh)^2 - 1}{h} V_t + 2\lambda + \mu e^{-\mu t}\left(E_0 - \tfrac{\lambda}{\mu}\right) = -2\mu V_t + 2\lambda + \mu e^{-\mu t}\left(E_0 - \tfrac{\lambda}{\mu}\right)$$
$$\Rightarrow V_t = e^{-2\mu t}(V_0 - E_0) + e^{-\mu t}\left(E_0 - \tfrac{\lambda}{\mu}\right) + \tfrac{\lambda}{\mu}$$

Appendix B. An alternative discrete-time model

At the core of the proposed predictive framework in this manuscript, a queuing model is employed to describe the stochastic occupancy change of a parking facility. The underlying queuing model can be any reasonable model. The main manuscript has demonstrated the proposed framework with the well-established continuous-time Markov M\M\C\C queue. This appendix will further discuss an alternative discrete-time model that focuses on the occupancy at a series of discrete time points with time interval Δt.

B.1. The discrete-time model

Instead of Poisson arrival and departure, both processes can be reasonably modeled as binomial distributions. This could potentially allow us to incorporate the parking-searching process into our framework, and could potentially be more computationally efficient. For arrivals, the probability that an approaching vehicle would select a given parking facility can be treated as a function of traffic and parking facility characteristics (and possibly driver characteristics). This approach, however, would require traffic flow data, which was not readily available for this study. On the other hand, the number of departure from a parking facility during Δt can be calculated from the number of parked vehicles and a leaving probability p. In view of data availability, we will demonstrate a discrete model with Poisson arrival and binomial departure in this appendix.

The number of arrivals during Δt is assumed to be Poisson with average λ. The probability of k vehicles arriving during time interval $[t,\ t+\Delta t)$, denoted as $P_a(k,\lambda)$ can be written as:

$$P_a(k,\lambda) = P(A(t) = k) = \begin{cases} \frac{\lambda^k}{k!}e^{-\lambda} & if\, k \in \{0\} \cup N \\ 0 & else \end{cases} \quad \text{(B.1)}$$

where $A(t)$ is the number of arrival during $[t,\ t+\Delta t]$. Note that not all arriving vehicles are guaranteed to find a parking spot when the parking lot becomes full. In this case, we assume these vehicles are blocked and will leave for an alternative parking lot immediately.

Let $D(t)$ denote the number of departures during time interval $[t,\ t+\Delta t)$; n the occupancy of the parking lot at time t; p the leaving probability of each vehicle in the parking lot during time interval Δt. The probability of k vehicles leaving in Δt, given occupancy n and leaving probability p can then be expressed as:

$$P_d(k,n,p) = P(D(t) = k) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k} & if\, k = 0, 1, \ldots, n \\ 0 & else \end{cases} \quad \text{(B.2)}$$

It is worth noticing that the average parking time (in terms of Δt) is equal to $1/p$. It is assumed that each vehicle's parking time is greater than Δt. Therefore, Δt should be set to a small enough value. In fact, when occupancy is sufficiently large and the leaving probability sufficiently small, the binomial distribution is a close approximation to Poisson distribution. This is also a well-known conclusion in probability theory.

With the above assumptions, the Markov matrix P of this alternative discrete-time model can be derived. Let $P_{i,j}$ denote the transition probability from occupancy $= i$ to j in time interval $[t, t+\Delta t)$.

$$P_{i,j} = \sum_{k=0}^{i} P_d(k, i, p) \times P_a(k+j-i, \lambda) \quad i, j = 0, 1, 2 \ldots\ldots, C-1 \tag{B.3a}$$

$$P_{i,C} = 1 - \sum_{k=0}^{C-1} P_{i,k} \quad i = 0, 1, 2, \ldots\ldots, C \tag{B.3b}$$

Note that arrival and departure as events can occur at any time and not necessarily at discrete time points. Rather, the discrete-time model assumes that the arrival and leaving probabilities can only change at discrete time points.

If the parking facility is under-saturated, similar results on time-dependent expectation and variance can be derived for the discrete model:

$$E_m = (1-p)^m \left(E_0 - \frac{\lambda}{p} \right) + \frac{\lambda}{p} \tag{B.4}$$

$$V_m = (1-p)^{2m} (V_0 - E_0) + (1-p)^m \left(E_0 - \frac{\lambda}{p} \right) + \frac{\lambda}{p} \tag{B.5}$$

When Eq. (1) from the continuous-time model is adopted in the regression method, the resulting expectation curve would be the same as that of the Eq. (B.4) from discrete-time model, although the estimated arrival and departure rates will not be the same. It is trivial to prove that from the same data set, the discrete-time model will lead to a higher estimated arrival rate and a lower estimated average parking time. But the fact that the expectation curves from the two models are the same means that the occupancy prediction will not be different.

B.2. Computational efficiency of transition probability matrix evaluation

Computational efficiency becomes relevant when it is necessary to evaluate the transition probability matrix. This not only occurs in parameter estimation when a facility is over-saturated, but also in probabilistic prediction of future occupancy regardless of whether a facility is over- or under-saturated. In a smart parking management scenario, the former is likely carried out by a server while the latter on the client's end to reduce communication cost.

In the continuous-time model, calculating the transition probability matrix involves evaluation of a matrix exponential operator $e^X := \sum_{k=0}^{\infty} X^k/k!$ (Coddington and Levinson, 1955), which is non-trivial. Moler and Van Loan (2003) provided an extensive review of various approaches to approximating matrix exponential. The applicability of the approaches depends on the matrix properties such as its 2-norm and eigenvalues. Different approaches also vary in reliability, stability, and accuracy. Moler and Van Loan (2003) identified the scaling-and-squaring method a potential "best" method as it is regarded generally applicable and has several reliable implementations. The computational complexity of the method is about $O((m+1/3)n^3)$ floating point operations,[3] where m is an integer ranging from 1 to about 15 depending on the 2-norm of the matrix and the desired error bound and n is the size of the square matrix. In our application of modeling parking process, n is the capacity of the parking facility plus one. Furthermore, note that in the continuous-time Markov model, the arrival and departure rates can also vary with time. In this case, solving for $P(t)$ becomes a linear ordinary differential equation with time-varying coefficients whose closed-form solution does not exist (Blanes et al., 2009).

For the alternative discrete-time model, the computational complexity of evaluating the transition probability matrix from Eqs. (B.1)–(B.3) is dominated by Eq. (B.3). In Eqs. (B.1) and (B.2), the exponentials only need to be calculated once; and the factorials and binomial coefficients can be calculated iteratively. For example, to compute $P_a(k,\lambda)$ $k = 0, 1, 2, \ldots C$ in formula (B.1), we only need to compute $e^{-\lambda}$ once when calculating $P_a(0,\lambda)$. We then apply $P_a(k+1, \lambda) = \frac{\lambda}{k+1} P_a(k, \lambda)$ to compute all the following probabilities iteratively. The total complexity of Eq. (1) is $O(C)$ plus the computation of $e^{-\lambda}$. Similarly, the complexity ofEq. (B.2) is also $O(C)$. On the other hand, it is not difficult to see that Eq. (B.3) requires $O(C^3/3)$ floating point operations. Therefore, the complexity of evaluating Eqs. (1)–(3) is $O(C^3/3)$, which is lower than the $O((m+1/3)C^3)$ from the continuous-time model. This shows that the discrete-time model is potentially more computationally efficient than the continuous-time model. The actual computer time needed would depend on the implementation of both approaches and the values of model parameters.

We performed numerical experiments on a DELL Precision T3610 workstation with 2.6 GHz CPU and 32GB RAM running Windows 8.1 Pro to compare the actual computation time of the transition probability matrix using both models. We implemented the discrete-time model in Matlab; for the continuous-time model, we adopted the Matlab built-in function expm() implementing the scaling-and-squaring method with Padé approximation. Our experiment results show that the discrete-time model performs better when the capacity is under 50 while the continuous-time model runs faster when the capacity

[3] As defined in Moler and Van Loan (2003), one floating-point computation "involves one floating point multiplication, one floating point addition, a few subscript and index calculations, and a few storage references".

is higher than 70. When the values of λ and p change, the computation time also changes. If the prediction is carried out by a device at a client's end (i.e., a typical smartphone), one possible approach to boost computational time is to reduce the number of significant digits by aggregating parking occupancy into different bins (for example a bin could consist 10 spots). This way, the size of the transition probability matrix is also reduced by the same factor and the discrete-model can be applied for faster computation for smaller matrices.

References

Anderson, S.P., de Palma, A., 2004. The economics of pricing parking. J. Urban Econ. 55 (1), 1–20.
Arnott, R., de Palma, A., Lindsey, R., 1991. A temporal and spatial equilibrium analysis of commuter parking. J. Public Econ. 45 (3), 301–335.
Arnott, R., Rowse, J., 1999. Modeling parking. J. Urban Econ. 45 (1), 97–124.
Ayala, D., Wolfson, O., Xu, B., Dasgupta, B., Lin, J., 2011. Parking slot assignment games. In: Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems.
Ayala, D., Wolfson, O., Xu, B., DasGupta, B., Lin, J., 2012. Pricing of parking for congestion reduction. In: Proceedings of the 20th International Conference on Advances in Geographic Information Systems, pp. 43–51.
Benenson, I., Martens, K., Birfir, S., 2008. PARKAGENT: an agent-based model of parking in the city. Comput. Env. Urban Syst. 32 (6), 431–439.
BestParking, 2016. BestParking. http://www.bestparking.com/. Accessed Jul 20. 2016.
Bifulco, G.N., 1993. A stochastic user equilibrium assignment model for the evaluation of parking policies. Eur. J. Operat. Res. 71 (2), 269–287.
Blanes, S., Casas, F., Oteo, J.A., Ros, J., 2009. The Magnus expansion and some of its applications. Phys. Rep. 470 (5), 151–238.
Boehlé, J.L., Rothkrantz, L.J.M., & Wezel, M.V. 2008. CBPRS: a city based parking and routing system.
Boyles, S.D., Tang, S., Unnikrishnan, A., 2015. Parking search equilibrium on a network. Transport. Res. Part B 81, 390–409.
Burns, M.R., Faurot, D.J., 1992. An econometric forecasting model of revenues from urban parking facilities. J. Econ. Bus. 44 (2), 143–150.
Caicedo, F., 2009. The use of space availability information in "PARC" systems to reduce search times in parking facilities. Transport. Res. Part C 17 (1), 56–68.
Caicedo, F., Blazquez, C., Miranda, P., 2012. Prediction of parking space availability in real time. Expert Syst. Appl. 39 (8), 7281–7290.
Caicedo, F., Robuste, F., Lopez-Pita, A., 2006. Parking management and modeling of car park patron behavior in underground facilities. Transp. Res. Rec. 1956, 60–67.
Caliskan, M., Barthels, A., Scheuermann, B., Mauve, M., 2007. Predicting parking lot occupancy in vehicular ad hoc networks. IEEE 65th Conference on Vehicular Technology.
Chaniotakis, E., Pel, A.J., 2015. Drivers' parking location choice under uncertain parking availability and search times: a stated preference experiment. Transport. Res. Part A 82, 228–239.
Chen, Z., Yin, Y., He, F., Lin, J.L., 2015. Parking reservation for managing downtown curbside parking. Transp. Res. Rec. 2498, 12–18.
Coddington, E., Levinson, N., 1955. Theory of Ordinary Differential Equations. Tata McGraw-Hill Education.
Delot, T., Cenerario, N., Ilarri, S., Lecomte, S., 2009. A cooperative reservation protocol for parking spaces in vehicular ad hoc networks. In: Proceedings of the 6th International Conference on Mobile Technology, Application & Systems, September. ACM, p. 30.
Dunning, A.E., 2006. Method and system for projecting dynamic parking availability based on an ongoing survey for remote lots with high demand. United States Patent 7049979.
Firebase, 2015. SF parking open data sets - firebase. Retrieved May 1, 2015, from https://www.firebase.com/docs/open-data/parking.html
Giuffrè, T., Siniscalchi, S.M., Tesoriere, G., 2012. A novel architecture of parking management for smart cities. Procedia – Social Behav. Sci. 53, 16–28.
Guo, L., Huang, S., Zhuang, J., Sadek, A., 2013. Modeling parking behavior under uncertainty: a static game theoretic versus a sequential neo-additive capacity modeling approach. Networks Spatial Econ. 13 (3), 327–350.
Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I., 2009. The WEKA data mining software: an update. SIGKDD Explorat. 11 (1).
He, F., Yin, Y., Chen, Z., Zhou, J., 2015. Pricing of parking games with atomic players. Transport. Res. Part B 73, 1–12.
Ibeas, A., Dell'Olio, L., Bordagaray, M., Ortúzar, J., de, D., 2014. Modelling parking choices considering user heterogeneity. Transport. Res. Part A 70, 41–49.
IBM, 2011. Global parking survey: drivers share worldwide parking woes. http://www-03.ibm.com/press/us/en/pressrelease/35515.wss. Accessed Jan. 28, 2016.
IBM, 2016. IBM SPSS - IBM analytics. http://www.ibm.com/analytics/us/en/technology/spss/spss.html. Accessed Jul. 30 2016.
IPI, 2015. 2015 emerging trends in parking, report on a survey conducted by the International Parking Institute. http://www.parking.org/wp-content/uploads/2016/02/TPP-2015-09-2015-Emerging-Trends-in-Parking.pdf. Accessed Feb. 15, 2016.
Ji, Y., Tang, D., Blythe, P., Guo, W., Wang, W., 2015. Short-term forecasting of available parking space using wavelet neural network model. IET Intell. Transport Syst. 9 (2), 202–209.
Kokolaki, E., Karaliopoulos, M., Stavrakakis, I., 2012. Leveraging information in parking assistance systems. IEEE Trans. Veh. Technol. 62 (9), 4309–4317.
Kolda, T.G., Lewis, R.M., Torczon, V., 2003. Optimization by direct search: new perspectives on some classical and modern methods. SIAM Rev. 45 (3), 385–482.
Liu, S., Guan, H., Yan, H., Yin, H., 2010. Unoccupied parking space prediction of chaotic time series. In: Proceedings of the 10th International Conference of Chinese Transportation Professionals, Beijing, pp. 2122–2131.
Liu, W., Yang, H., Yin, Y., 2014. Expirable parking reservations for managing morning commute with parking space constraints. Transport. Res. Part C 44, 185–201.
Lu, R., Lin, X., Zhu, H., Shen, X., 2009. SPARK: a new VANET-based smart parking scheme for large parking lots. In: INFOCOM 2009, IEEE. IEEE, pp. 1413–1421.
Mackowski, D., Bai, Y., Ouyang, Y., 2015. Parking space management via dynamic performance-based pricing. Transport. Res. Procedia 7, 170–191.
Mcguiness, B.E., Mcneil, S., 1991. Statistical models to predict commercial- and parking-space occupancy. J. Urban Plann. Dev. 117 (4), 129–139.
Millard-Ball, A., Weinberger, R.R., Hampshire, R.C., 2014. Is the curb 80% full or 20% empty? Assessing the impacts of San Francisco's parking pricing experiment. Transport. Res. Part A 63, 76–92.
Moler, C., Van Loan, C., 2003. Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. SIAM Rev. 45 (1), 3–49.
Murtagh, F., 1983. A survey of recent advances in hierarchical clustering algorithms. Comput. J. 26 (4), 354–359.
Pango, 2016. Pango. http://www.mypango.com/. Accessed Jul 20. 2016.
ParkMe Inc., 2016. ParkMe | the best way to find & compare cheap parking rates. https://www.parkme.com/how-it-works. Accessed Jul 20. 2016.
ParkMobile, 2016. Parkmobile _ parking made simple. http://us.parkmobile.com/#how-it-works. Accessed Jul. 20, 2016.
Parkopedia, 2016. Find parking lots, parking garages, parking meters, private garages - reserve parking - Parkopedia. http://en.parkopedia.com/. Accessed Jul 20. 2016.
ParkWhiz, 2016. ParkWhiz_ guaranteed parking _ find and book parking anywhere. http://www.parkwhiz.com/. Accessed Jul 20. 2016.
Portilla, A.I., Oreña, B.A., Berodia, J.L., Díaz, F.J., 2009. Using $M/M/\infty$ queueing model in on-street parking maneuvers. J. Transp. Eng. 135 (8), 527–535.
Qian, Z.(Sean), Rajagopal, R., 2013. Optimal parking pricing in general networks with provision of occupancy information. Procedia – Social Behav. Sci. 80, 779–805.
Qian, Z.S., Rajagopal, R., 2014. Optimal dynamic parking pricing for morning commute considering expected cruising time. Transport. Res. Part C 48, 468–490.
Rajabioun, T., Ioannou, P., 2015. On-street and off-street parking availability prediction using multivariate spatiotemporal models. IEEE Trans. Intell. Transp. Syst. 16 (5), 2913–2924.

Ross, S.M., 2014. Introduction to Probability Models. Academic Press.
Schlote, A., King, C., Crisostomi, E., Shorten, R., 2014. Delay-tolerant stochastic algorithms for parking space assignment. IEEE Trans. Intell. Transp. Syst. 15 (5), 1922–1935.
Shin, J.H., Jun, H.B., 2014. A study on smart parking guidance algorithm. Transport. Res. Part C 44, 299–317.
Shoup, D.C., 2006. Cruising for parking. Transport Policy 13 (November (6)), 479–486 2006.
Spothero, 2016. SpotHero - easy parking for all _ reserve parking now & save. https://spothero.com/. Accessed Jul 20. 2016.
Sun, D.J., Ni, X.Y., Zhang, L.H., 2016. A discriminated release strategy for parking variable message sign display problem using agent-based simulation. IEEE Trans. Intell. Transp. Syst. 17 (1), 38–47.
Tan, P., Steinbach, M., Kumar, V., 2005. Introduction to Data Mining, first ed. Pearson Addison Wesley, Boston.
Tamrazian, A., Qian, Z.(Sean), Rajagopal, R., 2015. Where is my parking spot? On-line and off-line prediction of time-varying parking occupancy. Transport. Res. Record 2489, 77–85.
Teodorović, D., Lučić, P., 2006. Intelligent parking systems. Eur. J. Operat. Res. 175 (3), 1666–1681.
Thompson, R.G., Takada, K., Kobayakawa, S., 2001. Optimisation of parking guidance and information systems display configurations. Transport. Res. Part C 9 (1), 69–85.
Tiedemann, T., Vögele, T., Krell, M.M., Metzen, J.H., Kirchner, F., 2015. Concept of a data thread based parking space occupancy prediction in a berlin pilot region. Workshops at the Twenty-Ninth AAAI Conference on Artificial Intelligence.
Van Der Waerden, P., Timmermans, H., Barzeele, P., 2011. Car drivers' preferences regarding location and contents of parking guidance systems stated choice approach. Transp. Res. Rec. 2245, 63–69.
Vickrey, W., 1994. Statement to the joint committee on Washington, DC metropolitan problems. J. Urban Econ. 36 (1), 42–65.
Vlahogianni, E.I., Kepaptsoglou, K., Tsetsos, V., Karlaftis, M.G., 2016. A real-time parking prediction system for smart cities. J. Intell. Transport. Syst. 20 (2), 192–204.
Waraich, R.A., Axhausen, K., 2012. An agent-based parking choice model. In: paper presented at the 91st Annual Meeting of the Transportation Research Board, vol. 2319, pp. 39–46.
Waterson, B.J., Hounsell, N.B., Chatterjee, K., 2001. Quantifying the potential savings in travel time resulting from parking guidance systems - a simulation case study. J. Oper. Res. Soc. 52 (10), 1067–1077.
Xiao, L., Liu, T., Huang, H., 2016. On the morning commute problem with carpooling behavior under parking space constraint. Transport. Res. Part B 91, 383–407.
Yang, H., Liu, W., Wang, X., Zhang, X., 2013. On the morning commute problem with bottleneck congestion and parking space constraints. Transport. Res. Part B 58, 106–118.
Ziat, A., Leroy, B., Baskiotis, N., Denoyer, L., 2016. Joint prediction of road-traffic and parking occupancy over a city with representation learning. In: Intelligent Transportation Systems (ITSC), 2016 IEEE 19th International Conference on. IEEE, pp. 725–730.
Zhang, X., Van Wee, G.P., 2011. Efficiency comparison of various parking charge schemes considering daily travel cost in a linear city. Eur. J. Transport Infrastruct. Res. 11 (2), 234–255.