Moment-Based Probabilistic Prediction of Bike Availability for Bike-Sharing Systems

Cheng Feng(✉), Jane Hillston, and Daniël Reijsbergen

LFCS, School of Informatics, University of Edinburgh, Scotland, UK
s1109873@sms.ed.ac.uk, jane.hillston@ed.ac.uk, dreijsbe@inf.ed.ac.uk

Abstract. We study the problem of future bike availability prediction of a bike station through the moment analysis of a PCTMC model with time-dependent rates. Given a target station for prediction, the moments of the number of available bikes in the station at a future time can be derived by a set of moment equations with an initial set-up given by the snapshot of the current state of all stations in the system. A directed contribution graph with contribution propagation method is proposed to prune the PCTMC to make it only contain stations which have significant contribution to the journey flows to the target station. The underlying probability distribution of the available number of bikes is reconstructed through the maximum entropy approach based on the derived moments. The model is parametrized using historical data from Santander Cycles, the bike-sharing system in London. In the experiments, we show our model outperforms the classic time-inhomogeneous queueing model on several performance metrics for bike availability prediction.

Keywords: Availability prediction · PCTMC models · Moment analysis · Maximum entropy reconstruction

1 Introduction

In recent years, we have seen significant growth of bike-sharing programs all over the world [1]. Public bike-sharing systems have been launched in many major cities such as London, Paris, and Vienna. Indeed, they have become an important part of urban transportation which provides improved connectivity to other modes of public transit. The concept of bike-sharing systems is rather simple: the system consists of a number of bike stations distributed over a geographic area (city). Each station is equipped with a limited number of bike slots in which public bikes can be parked. When users arrive at a station, they pick up a bike, use it for a while, and then return it to another station of their choice.

With the increasing popularity of the smart transport theme, there has been great interest from the research community in the intelligent management of bike-sharing systems. Topics include, but are not limited to, policy design [2,3], intelligent bike redistribution [4–6], and user journey planning [7,8]. The focus of this paper is on the probabilistic prediction of the number of available bikes in

G. Agha and B. Van Houdt (Eds.): QEST 2016, LNCS 9826, pp. 139–155, 2016.
DOI: 10.1007/978-3-319-43425-4_9

stations. Having a predictive model is of vital interest to both the user and the system administrator. The user can use it to identify likely origin/destination stations for which a trip can be successfully made. System administrators can use the model to undertake service level agreement checking, and plan bike redistribution for stations which are likely to break the service level requirement.

In this paper we present a novel moment-based prediction model that can provide probabilistic forecasts for the number of available bikes in a bike station. By representing the bike-sharing system as a Population Continuous Time Markov Chain (PCTMC) with time-dependent rates, our model is explanatory as the dynamics of the system is explicitly given. Gast *et al.* [8] show the benefits of predicting (*forecasting*) the entire probability distributions of possible bike availabilities in a station, compared with previous models that were only able to produce point estimates, often using time-series-based techniques [7,9,10]. However, unlike [8], in which all the considered forecasting methods worked on the level of isolated stations, our model also captures the journey dynamics between stations. Guenther and Bradley [11] also provide a inhomogeneous-time PCTMC model with time-dependent rates for bike availability prediction, however there are several key differences between that model and ours. Firstly, our model provides the full probability distribution of the number of available bikes in a station whereas their model only provides a point estimate. Secondly, we use a model reduction method to prune our PCTMC such that the significant journey dynamics with respect to the target station are guaranteed to be preserved. However, their model aggregates stations which are spatially close, assuming that they have similar journey durations to the target station, which causes the information about the emptiness and fullness of stations to be lost.

We summarize the contribution of our paper as follows. Firstly, a novel PCTMC model with time-dependent rates is presented to successfully capture the journey dynamics between bike stations. Secondly, we propose a novel model reduction technique to prune the PCTMC model based on the directed contribution graph with a contribution propagation method for a given target station for bike availability prediction. Finally, we reconstruct the underlying probability distribution of the number of available bikes in the target station using the maximum entropy principle based on a few moments generated from fluid approximation of the PCTMC, and show that the model has a better performance on a set of metrics for bike availability prediction compared with the classic Markov single-station queueing model.

The rest of this paper is structured as follows. We briefly introduce the concepts of PCTMC with time-dependent rates in the next section. Section 3 gives the introduction of the classic Markov queueing model for bike availability prediction. In Sect. 4, we present our PCTMC model for the bike-sharing scenario. In the next section we show how to reconstruct the probability distribution of number of available bikes using the maximum entropy approach. Section 6 presents the experimental results of our model on the London bike-sharing system compared with the classic Markov queueing model. Finally, Sect. 7 discusses possible extensions of our model and draws final conclusions.

2 PCTMC with Time-Dependent Rates

A PCTMC is a stochastic process which consists of a number of distinct agent populations and a set of transition classes. The state of a PCTMC is captured by an integer vector counting the number of each agent type. The model evolves with the firing of transitions. When a transition fires, one or more agent populations are updated. Each transition is associated with a rate function, which assigns a rate governed by an exponential distribution to the transition based on the current state of the PCTMC. In this paper, we specifically consider time-inhomogeneous PCTMCs, in which transition rates can also be time-dependent. Specifically, a PCTMC with time-dependent rates can be expressed as a tuple $\mathcal{P} = (\mathbf{X}(t), \mathcal{T}, \mathbf{X_0})$:

- $\mathbf{X}(t) = (X_1(t), ..., X_n(t)) \in \mathbb{Z}^n_{\geq 0}$ is an integer vector with the ith $(1 \leq i \leq n)$ component representing the current number of an agent type S_i.
- $\mathcal{T} = \{\tau_1, ..., \tau_m\}$ is the set of transition classes, of the form $\tau = (r_\tau(\mathbf{X}, t), \mathbf{d}_\tau)$, where:
  1. $r_\tau(\mathbf{X}, t) \in \mathbb{R} \geq 0$ is a time-dependent rate function, associating with each transition the rate of an exponential distribution, depending on the state of the PCTMC $\mathbf{X}$ as well as the current time t.
  2. $\mathbf{d}_\tau \in \mathbb{Z}^n$ is the update vector which gives the net change for each element of $\mathbf{X}$ caused by transition τ.
- $\mathbf{X_0} \in \mathbb{Z}^n_{\geq 0}$ is the initial state of the model.

Transition rules can be easily expressed in the chemical reaction style, as

$$\underline{\ell}_1 S_1 + \ldots + \underline{\ell}_n S_n \longrightarrow_\tau \overline{\ell}_n S_1 + \ldots + \overline{\ell}_n S_n \quad \text{at rate}\, r_\tau(\mathbf{X}, t)$$

where the net change of agents of type S_i due to transition τ is given by $d^i_\tau = \overline{\ell}_i - \underline{\ell}_i$ $(1 \leq i \leq n)$, and the transition rate is $\begin{cases} r_\tau(\mathbf{X}, t) & \text{if } X_i \geq \underline{\ell}_i \quad \forall i = 1, 2, \ldots, n \\ 0 & \text{otherwise.} \end{cases}$

As the state space of PCTMC models is often very large or even infinite, numerical techniques traditionally used for performance analysis, based on a Markovian approach, are entirely infeasible. Stochastic simulation is feasible, but deriving useful metrics such as mean, variance, probability distribution of populations often requires a large number of simulation runs, thus making this approach extremely costly in terms of computational resources, particularly when estimating full probability distributions over large state spaces. In this paper, we will adopt a much more computationally efficient approach to analyse the PCTMC for the bike-sharing model. Specifically, we approximate the evolution of the moments of the underlying population-level stochastic process of a PCTMC model by the following set of ODEs [12]:

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[M(\mathbf{X}(t))] = \sum_{\tau \in \mathcal{T}} \mathbb{E}[(M(\mathbf{X}(t) + \mathbf{d}_\tau) - M(\mathbf{X}(t)))r_\tau(\mathbf{X}, t)] \tag{2.1}$$

where $M(\mathbf{X})$ denotes the moment to be calculated. For instance, by substituting $M(\mathbf{X})$ with X_i, ${X_i}^2$ and X_iX_j, we get the set of ODEs to describe the first moment, second moment and second-order joint moment respectively, of population variables in an arbitrary PCTMC model. The set of ODEs can be directly solved by numerical simulation as long as there is no transition rate in the PCTMC with non-linear polynomials. With time-dependent rates, the system becomes hybrid with discrete jumps of rates at some specific points of numerical simulation.

3 Markov Queueing Model

Before introducing our model, we first give the traditional Markov queueing model for bike stations which is going to serve as our comparator.

The most straightforward way to evaluate the behaviour of a station is to analyse it in isolation. In this case, a station can be modelled as a time-inhomogeneous Markov queue $M/M/1/k_i$, illustrated in Fig. 1.

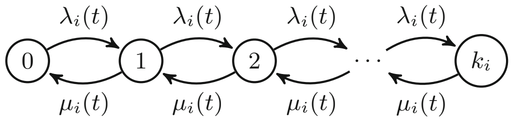

Fig. 1. The time-inhomogeneous Markov queue for station i

Specifically, k_i denotes the capacity of a station i, $\lambda_i(t)$ and $\mu_i(t)$ are the time-dependent bike arrival and pickup rates of station i at time t of a day. Usually, the time of a day is split into n even slots.

Then, using the transition rate matrix for station i: $Q(\lambda_i(t), \mu_i(t))$, where

$$Q(\lambda, \mu) = \begin{pmatrix} -\mu & \mu & & & \\ \lambda & -(\mu+\lambda) & \mu & & \\ & \ddots & \ddots & \ddots & \\ & & \lambda & -(\mu+\lambda) & \mu \\ & & & \lambda & -\lambda \end{pmatrix},$$

one can predict the probability that there are y bikes in station i at time $t+h$ given the station has x bikes at time t, by the following equation:

$$\Pr(y \mid x, t, h) = \exp\left(\int_0^h Q(t+s)\mathrm{d}s\right)_{x,y}$$

where $\exp(M)_{x,y}$ is the element at row x and column y of the matrix exponential of M. Such a model has been used to make bike availability or station inventory level predictions in several papers in the literature (e.g. [6,8,13]).

Two assumptions are made in this model. First, the bike arrivals and pickups at stations form Poisson processes. Second, the state of a particular station does not depend on the state of the others. The first assumption is successfully validated for busy stations in [8], using historical data from the Velib bike-sharing system in Paris. However, we conjecture that the second assumption is generally not true in practice. For example, when a station is empty, no bikes can depart from it, therefore the arrival rate at other stations should be reduced. Hence, we seek a more realistic model, which captures the journey dynamics between stations.

4 PCTMC of Bike-Sharing Model

4.1 A Naive PCTMC Model

To faithfully represent the journey dynamics between bike stations in a bike-sharing system with N stations, we first propose a naive PCTMC model which contains the following transitions:

$$\begin{aligned} &Bike_i \longrightarrow Slot_i + Journey^i_j@P_1 \quad \text{at } \mu_i(t)p^i_j(t) && \forall i,j \in (1,N) \\ &Journey^i_j@P_l \longrightarrow Journey^i_j@P_{l+1} \quad \text{at } (P^i_j/d^i_j)\,\#(Journey^i_j@P_l) \\ & && l \geq 1 \wedge l < P^i_j,\ \forall i,j \in (1,N) \\ &Journey^i_j@P_{P^i_j} + Slot_j \longrightarrow Bike_j \quad \text{at } (P^i_j/d^i_j)\,\#(Journey^i_j@P_{P^i_j}) && \forall i,j \in (1,N) \end{aligned}$$

where $Bike_i$, $Slot_i$ represent a bike and a slot agent in station i respectively; $Journey^i_j@P_l$ represents a bike agent which is currently on a journey from station i to station j at phase l. Note that since journey durations are generally not exponentially distributed, we fit the journey duration from station i to station j as an Erlang distribution with P^i_j phases each with rate P^i_j/d^i_j, where d^i_j is the mean journey duration. $\mu_i(t)$ is the bike pickup rate in station i at time t, p^i_j is the probability that a journey will end at station j given that it started from station i at time t. $\#(S)$ denotes the population of an agent type S.

Obviously, the above model is not scalable. Since the total number of bike stations N is usually very large (for example there are around 750 bike stations in London), it is computationally infeasible to analyse a model which captures the full set of bike stations. Fortunately, since we are only interested in the prediction of bike availability of a single target station at a time, we only need to model stations which have a significant contribution to the journey flows to the target station (knowing the state of a station which has a very small contribution to the journey flows to the target station will have negligible impact on the accuracy of bike availability prediction for the target station). Thus, a directed contribution graph together with a contribution propagation method is proposed to automatically identify the set of stations which need to be modelled with respect to a given target station for bike availability prediction.

4.2 Directed Contribution Graph with Contribution Propagation

Here, we show how to derive a set of bike stations $\Theta(v)$ in which all stations have a significant contribution to the journey flows to a given target station $v \in (1, 2, \ldots, N)$ for bike availability prediction. Concretely, we first need a way to quantify the contribution of one station to the journey flows to another station. Specifically, we let C_{ij} denote the *contribution coefficient* of station j to station i which quantifies the contribution of station j to the journey flows to station i.

One station can contribute to the journey flows to another station both directly and indirectly. The definition of a *direct contribution coefficient* at time t is given by the following simple formula:

$$c_{ij}(t) = \lambda_i^j(t)/\lambda_i(t)$$

in which $\lambda_i^j(t)$ represents the bike arrival rate from station j to station i at time t and $\lambda_i(t) = \sum_j \lambda_i^j(t)$. Then, it is clear that $c_{ij}(t) \in [0, 1]$, $0 \leq \sum_{j \neq i} c_{ij}(t) \leq 1$.

With the definition of directed contribution coefficient, we can construct a directed contribution graph for the bike-sharing system at each time slot of a day. The definition of the directed contribution graph is given as follows (for convenience, we abbreviate $c_{ij}(t)$ to c_{ij}):

Definition 1. *For an arbitrary time t, the directed contribution graph for a bike-sharing system at time t is a graph in which nodes represent the stations in the system, and there is a weighted directed edge from node i to node j if $c_{ij} > 0$, and in this case the weight of the edge is c_{ij}. Thus, the direction of edges is the inverse of contribution flows.*

Figure 2 shows a sample directed contribution graph which consists of six bike stations.

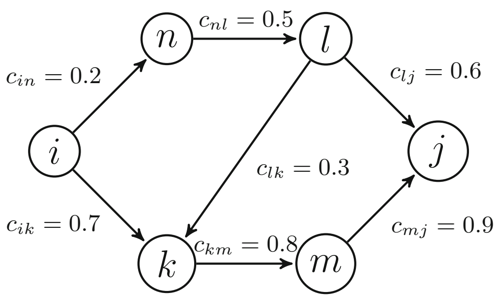

Fig. 2. An example directed contribution graph with six stations

For those stations which are not directly connected in the directed relation graph, by using a contribution propagation method, we can evaluate the *indirect* *contribution coefficient* of one station on the journey flows to another station. Specifically, the indirect contribution coefficient is quantified by a path dependent coefficient $c_{ij,\gamma}$, which is the product of the direct contribution coefficients along an acyclic path γ from node i to node j. Then, the contribution coefficient of station j to station i is characterized by the maximum of the path dependent coefficients:

$$c_{ij,\gamma} = \prod_{kl \in \gamma} c_{kl}$$

$$C_{ij} = \begin{cases} \max\limits_{all\ paths\ \gamma} c_{ij,\gamma} & \text{if there exists a path from node } i \text{ to node } j \\ 0, & \text{otherwise} \end{cases}$$

For example, according to Fig. 2, the contribution coefficient of station j to station i is $C_{ij} = c_{ik} \times c_{km} \times c_{mj} = 0.504$, since $c_{ik} \times c_{km} \times c_{mj} > c_{in} \times c_{nl} \times c_{lj} > c_{in} \times c_{nl} \times c_{lk} \times c_{km} \times c_{mj}$.

With the contribution coefficient, given a target station v, then for $i \in (1, 2, \ldots, N)$, we can infer:

$$\begin{aligned} &i \in \Theta(v) && \text{if } C_{vi} > \theta \\ &i \notin \Theta(v) && \text{if } C_{vi} \leq \theta \end{aligned}$$

where $\theta \in (0, 1)$ is threshold value which can be used to control the extent of model reduction. A point to note is that we choose to characterize contribution coefficients by the maximum instead of the sum of path dependent coefficients because we only want to model stations which have at least a significant (direct or indirect) journey flow to the target station. To model stations which have many small journey flows to the target station is costly but the impact is rather unpredictable. Moreover, the maximum of path dependent coefficients has another nice property that if $i \in \Theta(v)$ and $C_{vi} = c_{vi,\gamma}$, then for a station j which is on the path γ, it is certain that $C_{vj} > \theta$, thus $j \in \Theta(v)$. As a result, for all stations which have a significant journey flow to the target station, that journey flow will certainly be captured in the resulting reduced PCTMC. However, this property will not be preserved if we use the sum of path dependent coefficients. For example in Fig. 2, if we set $\theta = 0.55$, then $\sum_{\gamma} c_{ij,\gamma} > \theta$, thus station j is included in the reduced PCTMC. However, since $\sum_{\gamma} c_{il,\gamma} < \theta$, station l will not be included, thus $\sum_{\gamma} c_{ij,\gamma} < \theta$ will not be satisfied in the reduced PCTMC.

As an illustration of the extent of model reduction, Fig. 3 shows the empirical cumulative distribution function of contribution coefficients between all bike stations during all time slots (which is computed by journey data from the London Santander Bike-sharing system, with 20 min slot duration). It can be seen that more than 96 % stations can be excluded even if θ is set to the small value 0.01.

4.3 The Reduced PCTMC Model

Given a target station v and current time t, suppose we are interested in the number of bikes at the station at time $t + h$, then let $s = (s_1, s_2, \ldots, s_n)$ be the

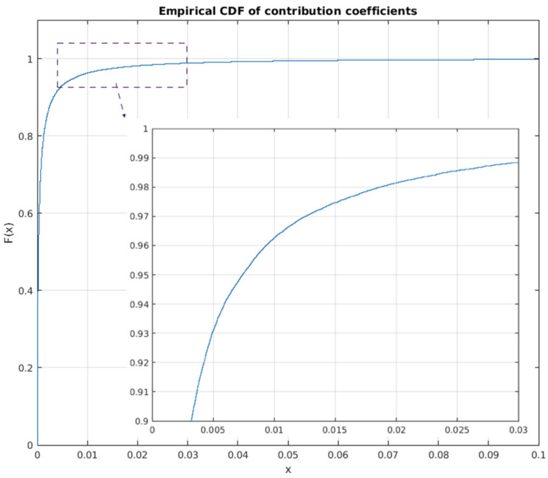

Fig. 3. The empirical cumulative distribution function of contribution coefficients (x is the value of contribution coefficients)

minimal set of time slots which covers $[t, t+h]$, we obtain $\Theta(v) = \Theta(v, s_1) \cup \Theta(v, s_2) \cup \ldots \cup \Theta(v, s_n) \cup v$, where $\Theta(v, s_i)$ is the set of bike stations which have significant contribution to the journey flows to the target station at time slot s_i.

Therefore, the PCTMC for the prediction of bike availability at station v at time $t+h$ can be represented as follows:

$$Bike_i \longrightarrow Slot_i \quad \text{at } \mu_i(t)\Big(1 - \sum_{j \notin \Theta(v) \vee c_{ji} \leq \theta} p_j^i(t)\Big) \qquad \forall i \in \Theta(v) \tag{4.1}$$

$$Slot_i \longrightarrow Bike_i \quad \text{at } \sum_{j \notin \Theta(v) \vee c_{ij} \leq \theta} \lambda_i^j(t) \qquad \forall i \in \Theta(v) \tag{4.2}$$

$$Bike_i \longrightarrow Slot_i + Journey_j^i@P_1 \quad \text{at } \mu_i(t) p_j^i(t) \quad \forall i, j \in \Theta(v) \wedge c_{ji} > \theta \tag{4.3}$$

$$Journey_j^i@P_l \longrightarrow Journey_j^i@P_{l+1} \quad \text{at } (P_j^i/d_j^i)\,\#(Journey_j^i@P_l)$$
$$l \geq 1 \wedge l < P_j^i, \forall i, j \in \Theta(v) \wedge c_{ji} > \theta \tag{4.4}$$

$$Slot_j + Journey_j^i@P_{P_j^i} \longrightarrow Bike_j \quad \text{at } (P_j^i/d_j^i)\,\#(Journey_j^i@P_{P_j^i})$$
$$\forall i, j \in \Theta(v) \wedge c_{ji} > \theta \tag{4.5}$$

$$Journey_j^i@P_{P_j^i} \longrightarrow \varnothing \quad \text{at } \mathbb{1}\big(Slot_j(t) = 0\big)(P_j^i/d_j^i)\,\#(Journey_j^i@P_{P_j^i})$$
$$\forall i, j \in \Theta(v) \wedge c_{ji} > \theta \tag{4.6}$$

where (4.1) represents a bike in station i is picked up for a journey to a station outside $\Theta(v)$ or a station to which the journey flow is negligible (the direct contribution coefficient $c_{ji} \leq \theta$ indicates that journey flow from i to j must not be a significant journey flow); (4.2) represents a bike is returned to station i from a station outside $\Theta(v)$ or a station from which the journey flow is negligible;

(4.3) represents a bike in station i is picked up for a journey to a station j inside $\Theta(v)$ and the journey flow is significant; (4.4), (4.5) represent progress and completion of the journey, respectively; (4.6) assumes a bike in transit from station i to station j will be returned to another station outside $\Theta(v)$ when there is no empty slot in station j, where $\mathbb{1}(Slot_j(t) = 0)$ is an indicator function which returns 1 when the number of empty slots at station j at time t is zero, otherwise returns 0.

Dealing with Indicator Function. Since we are going to numerically solve the PCTMC using moment ODEs as illustrated in Eq. (2.1), we can only access the moments of the number of empty slots at a station i at time t, denoted as u_i^m, during numerical simulation (here we let u_i^m denote $\mathbb{E}[\big(Slot_i(t)\big)^m]$, where m is the order of the moment), whereas the number of empty slots at station i at time t is a random variable. Thus, we propose a method to approximate the indicator function by a function of the moments u_i^m of the number of empty slots and the capacity of the station: $\mathbb{1}(Slot_i(t) = 0) \sim f(u_i^1, u_i^2, \ldots, u_i^m, k_i)$. Concretely, given the first m moments of the random variable $Slot_i(t)$, and the value domain $Slot_i(t) \in [0, 1, \ldots, k_i]$, we can approximate the probability distribution of $Slot_i(t)$ by a discrete distribution with finite support k_i. For example, if we only know the first moment of $Slot_i(t)$ (which is u_i^1), we can fit a binomial distribution $Slot_i(t) \sim Binomial(k_i, u_i^1/k_i)$ to the probability distribution of $Slot_i(t)$. In this case, we get $\Pr(Slot_i(t) = 0) = (1 - u_i^1/k_i)^{k_i}$. Furthermore, if we know the first two moments (u_i^1, u_i^2), then we can fit a beta-binomial distribution $Slot_i(t) \sim BetaBinomial(k_i, \alpha, \beta)$, where

$$\alpha = \frac{u_i^1 u_i^2 - k_i(u_i^1)^2}{k_i(u_i^1)^2 + k_i u_i^1 - k_i u_i^2 - (u_i^1)^2} \qquad \beta = \frac{(k_i - u_i^1)(k_i u_i^1 - u_i^2)}{k_i(u_i^1)^2 + k_i u_i^1 - k_i u_i^2 - (u_i^1)^2}$$

Thus, we get

$$\Pr(Slot_i(t) = 0) = \frac{B(\alpha, k_i + \beta)}{B(\alpha, \beta)}$$

where $B(a, b)$ is a beta function. Theoretically, with knowledge of more moments of $Slot_i(t)$, the estimation of $\Pr(Slot_i(t) = 0)$ will be more accurate. Finally, we let

$$\mathbb{1}(Slot_i(t) = 0) = \begin{cases} 1 & \text{if } \Pr(Slot_i(t) = 0) > p \\ 0 & \text{if } \Pr(Slot_i(t) = 0) \le p \end{cases}$$

where $\Pr(Slot_i(t) = 0) = f(u_i^1, u_i^2, \ldots, u_i^m, k_i)$, p is a threshold value beyond which we believe the number of empty slots in station i is zero. In general p should be set to a value close to 1. In our later experiments, we explicitly set $p = 0.9$.

Specifying the Initial State. Given a snapshot of the bike-sharing system at a time instant t which contains the following information[1]:

$$Bike_i(t), \ldots, Slot_i(t), \ldots, Journey^i(t, \Delta t), \ldots$$

[1] This information is actually recorded for the London bike-sharing system.

where $Bike_i(t)$ and $Slot_i(t)$ are the current number of available bikes and empty slots at a station i; $Journey^i(t, \Delta t)$ represents there is a bike currently en route from station i, and the journey started at time $t - \Delta t$. Then, for each $Journey^i(t, \Delta t)$, we use a random number to determine the destination of the journey, and the time Δt to determine the appropriate phase of the journey time. Thus we generate a random number α uniformly distributed in $(0, 1)$, and let $p^i_k(t - \Delta t), \forall k$ be the probability that the journey will end at station k given that the journey started from station i at time $t - \Delta t$. Then

$$Journey^i(t, \Delta t) = Journey^i_j(t, \Delta t) \text{ if } \alpha \geq \sum_{k=0}^{j-1} p^i_k(t - \Delta t) \text{ and } \alpha < \sum_{k=0}^{j} p^i_k(t - \Delta t).$$

Furthermore, we let

$$Journey^i_j(t, \Delta t) = Journey^i_j@P_l \text{ if } \Delta t \geq (l-1)d^i_j/P^i_j \text{ and } \Delta t < l \times d^i_j/P^i_j,$$

where $l \leq P^i_j$. Otherwise, if $l > P^i_j$, we let $Journey^i_j(t, \Delta t) = Journey^i_j@P_{P^i_j}$.

Solving the Moment ODEs. We derive the moment ODEs following Eq. (2.1) for the above PCTMC for the first m order of moments. Furthermore, using the correlation heuristics introduced in [14], we can make a further reduction on the size of the moment ODEs, utilizing the neighbourhood relation between agents in the above PCTMC. Specifically, we let $\mathbb{E}[(X_i)^{m_i}(X_j)^{m_j}] \approx \mathbb{E}[(X_i)^{m_i}]\mathbb{E}[(X_j)^{m_j}]$ if there does not exist a transition in the PCTMC in which both agent S_i and S_j are directly involved. Due to limited space, we refer to [14] for more detail of the reduction algorithm. The moment ODEs can be solved by numerical simulation using standard methods.

5 Reconstructing the Probability Distribution Using the Maximum Entropy Approach

From the moment analysis of the PCTMC for bike-sharing model, we gain the first m moments of the number of available bikes in the target station at the prediction time $t+h$, i.e. $\left(\left(Bike_v(t+h)\right)^1, \left(Bike_v(t+h)\right)^2, \ldots, \left(Bike_v(t+h)\right)^m\right)$, which we denote as $(u^1, u^2, \ldots, u^m)$ in the following. Our goal is to predict the probability that the station has a specific number of bikes at time $t + h$. This means the problem is to reveal $\Pr\left(Bike_v(t + h) = i \mid u^1, u^2, \ldots, u^m, k_v\right)$, where $i \in (1, 2, \ldots, k_v)$. Therefore, we need to reconstruct the entire probability distribution of the random variable $Bike_v(t + h)$ based on its first m moments. The corresponding distribution is generally not uniquely determined. Hence, to select a particular distribution, we apply the maximum entropy principle to minimize the amount of bias in the reconstruction process. In this way, we assume the least amount of prior information about the true distribution. Note that the maximum entropy approach has been successfully applied to reconstruct distributions based on moments in many areas, e.g. physics [15], stochastic chemical kinetics [16], and performance analysis [17].

5.1 Reconstruction Algorithm

Let X_v denote $Bike_v(t+h)$ for convenience, $\mathcal{G}$ be the set of all possible probability distributions for X_v. Then, based on the maximum entropy principle, the goal is to select a distribution g to maximize the entropy $H(g)$ over all distributions in $\mathcal{G}$. The problem can be denoted as follows:

$$\arg\max_{g \in \mathcal{G}} H(g) = \arg\max_{g \in \mathcal{G}} \big(- \sum_{x=0}^{k_v} g(x) \ln g(x) \big)$$

Furthermore, given $(u^1, u^2, \ldots, u^m)$, we know the following constraints should be satisfied:

$$\sum_{x=0}^{k_v} x^n g(x) = u^n, \quad n = 0, 1, \ldots, m$$

where $u^0 = 1$ to ensure that g is a probability distribution. Now, the problem becomes a constrained optimization program. Thus to perform the constrained maximization of the entropy, we introduce one Lagrange multiplier λ_n per moment constraint. We thus seek extrema of the Lagrangian functional:

$$L(g, \lambda) = - \sum_{x=0}^{k_v} g(x) \ln g(x) - \sum_{n=0}^{m} \lambda_n \big(\sum_{x=0}^{k_v} x^n g(x) - u^n \big)$$

Functional variation with respect to the unknown distribution function $g(x)$ yields:

$$\frac{\partial L}{\partial g(x)} = 0 \implies g(x) = \exp\left(-1 - \lambda_0 - \sum_{n=1}^{m} \lambda_n x^n \right)$$

Since $u^0 = 1$, we get

$$\sum_{x=0}^{k_v} \exp\left(-1 - \lambda_0 - \sum_{n=1}^{m} \lambda_n x^n \right) = 1.$$

Thus we can express λ_0 in terms of the remaining Lagrange multipliers

$$e^{1+\lambda_0} = \sum_{x=0}^{k_v} \exp\left(- \sum_{n=1}^{m} \lambda_n x^n \right) \equiv Z$$

Then, the general form of $g(x)$ can be given as follows:

$$g(x) = \frac{1}{Z} \exp\left(- \sum_{n=1}^{m} \lambda_n x^n \right)$$

Insert the preceding equation into the Lagrangian, we can then transform the problem into an unconstrained minimization problem of the following function with respect to variables $\lambda_1, \lambda_2, \ldots, \lambda_n$:

$$\varGamma(\lambda_1, \lambda_2, \ldots, \lambda_n) = \ln Z + \sum_{n=1}^{m} \lambda_n u^n$$

The convexity of the function Γ is proved in [15], which guarantees the existence of a unique solution. Thus, a close approximation $(\lambda_1^*, \lambda_2^*, \ldots, \lambda_n^*)$ of the true solution can be obtained by the classic gradient descent approach [18].

Thus, after finding $(\lambda_1^*, \lambda_2^*, \ldots, \lambda_n^*)$ through gradient descent, we can finally predict

$$\Pr\big(X_v = x\big) = \frac{\exp\Bigg(-\sum_{n=1}^{m}\lambda_n^* x^n\Bigg)}{\sum_{i=0}^{k_v}\exp\Bigg(-\sum_{n=1}^{m}\lambda_n^* i^n\Bigg)}, \quad \forall x \in (1, 2, \ldots, k_v).$$

6 Experiments

In this section, we test the time cost and accuracy of our prediction model in different cases and compare the accuracy of our model with the classic Markov queueing model. We use the historic journey data and bike availability data from January 2015 to March 2015 from the London Santander Cycles Hire scheme to train our PCTMC model as well as the Markov queueing model, and the data in April 2015 to test their prediction accuracy. As in [11], we fit the number of journey phases between stations using the HyperStar tool [19] command line interface. Specifically, we set the maximum value of P_j^i to 20 to make our model compact and also avoid overfitting. Moreover, for parameters estimation, we split a day into slots of 20 min duration. In our experiments, given the bike availability in a station at time t, we predict the probability distribution of the number of available bikes in that station at time $t + h$, where h is set to 10 min for short range prediction and 40 min for long range prediction.

The evaluation of our model is twofold. The first is accuracy, the second is efficiency. These two aspects are both influenced by the value of two important parameters, namely m, the highest order of moments being derived, and θ, the coefficient threshold for the identification of bike stations which have significant contribution to the journey flow to the target station. For higher values of m, the solution cost of our model becomes larger since more moment ODEs are derived, however the model should become more accurate due to more constraints in the probability distribution reconstruction based on the maximum entropy principle. For higher values of θ, more stations are excluded in the reduced PCTMC for a target station whereas the model accuracy can be potentially reduced. Thus, to observe the effects on these two parameters, we do experiments with values $m = 1, 2, 3$, $\theta = 0.01, 0.02, 0.03$.

6.1 Root Mean Square Error

For prediction accuracy, we first consider the classic criterion based on root mean square error (RMSE), a commonly used metric for evaluating point predictions (i.e., predictions that only state the expected number of bikes).

Table 1. The calculated RMSE on the prediction of the number of available bikes

	10 min	40 min	
Markov queueing model	1.52	3.03	
PCTMC with $\theta = 0.03$	1.49	2.81	$m = 1, 2, 3$
PCTMC with $\theta = 0.02$	1.49	2.81	$m = 1, 2, 3$
PCTMC with $\theta = 0.01$	1.48	2.79	$m = 1, 2, 3$

Table 1 compares the RMSE of the prediction results of our PCTMC model with the Markov queueing model. As can be seen, the PCTMC model outperforms the Markov queueing model in both prediction ranges. Especially in the long range, a considerable improvement is observed. For the PCTMC models, smaller values of θ only reduce the RMSE slightly. This means capturing less significant journey flows will have little impact on the prediction accuracy. Moreover, we find that the derived highest moments have almost no impact on the RMSE. This is obvious since the expected number of available bikes is only decided by the first moment.

6.2 Probability of Making a Right Recommendation

Predicting the expected number of available bikes is important for system administrators when they want to decide how to redistribute bikes in the system. However, a user is interested in whether there is a bike in the target station when she wants to pick up a bike from there, or whether there is a free slot in the target station when she wants to return a bike to that station. We are specifically interested in being able to make correct recommendations for the queries "Will there be a bike?" and "Will there be a slot?"[2] to measure the accuracy of our model. Specifically, for the "Will there be a bike?" query, we respond "Yes" if the predicted probability of that station having more than one bike is greater than 0.8, and respond "No" if the predicted probability of that station having more than one bike is less than 0.8. As is argued in [8], the root mean square error is not an appropriate evaluation metric in this setting. After all, we need a prediction of the probability of the recommendation being correct rather than just a point estimate of the number of available bikes/slots. Instead, a suitable evaluation scheme is proposed in [8] that ensures that the best prediction algorithm can always be expected to obtain the highest score. Such a scheme is called a *proper scoring rule*. For the setting described above, the following scoring rule is proper:

$$\text{Score} = \begin{cases} 1 & \text{if } \Pr(X_v > 0) > 0.8 \wedge x_v > 0 \\ -4 & \text{if } \Pr(X_v > 0) > 0.8 \wedge x_v = 0 \\ 1 & \text{if } \Pr(X_v > 0) < 0.8 \wedge x_v = 0 \\ -\frac{1}{4} & \text{if } \Pr(X_v > 0) < 0.8 \wedge x_v > 0 \end{cases}$$

[2] These queries can be readily extended to "Will there be n bikes?" and "Will there be n slots?".

Table 2. Average score of making a recommendation to the "Will there be a bike?" query with 95 % confidence interval

	10 min	40 min	
Markov queueing model	0.9 ± 0.05	0.87 ± 0.06	
PCTMC with $\theta = 0.03$	0.91 ± 0.04	0.89 ± 0.05	$m = 2$
	0.92 ± 0.04	0.91 ± 0.04	$m = 3$
PCTMC with $\theta = 0.02$	0.91 ± 0.04	0.89 ± 0.05	$m = 2$
	0.92 ± 0.04	0.91 ± 0.04	$m = 3$
PCTMC with $\theta = 0.01$	0.92 ± 0.04	0.89 ± 0.05	$m = 2$
	0.93 ± 0.04	0.91 ± 0.04	$m = 3$

Table 3. Average score of making a recommendation to the "Will there be a slot?" query with 95 % confidence interval

	10 min	40 min	
Markov queueing model	0.91 ± 0.04	0.88 ± 0.05	
PCTMC with $\theta = 0.03$	0.91 ± 0.04	0.9 ± 0.05	$m = 2$
	0.92 ± 0.04	0.91 ± 0.04	$m = 3$
PCTMC with $\theta = 0.02$	0.91 ± 0.04	0.9 ± 0.05	$m = 2$
	0.92 ± 0.04	0.91 ± 0.04	$m = 3$
PCTMC with $\theta = 0.01$	0.92 ± 0.04	0.91 ± 0.05	$m = 2$
	0.93 ± 0.04	0.92 ± 0.04	$m = 3$

Note that incorrect predictions need to be penalised by a negative score for the rule to be proper. The evaluation of recommendations to the "Will there be a slot?" query follows a similar pattern. Tables 2 and 3 show the experimental results for different models and parameters. Note that the PCTMC model with $m = 1$ is excluded since at least two moments are needed to make a meaningful reconstruction of the probability distribution. As can be seen from the tables, the PCTMC model clearly has a better performance in making such recommendations. Moreover, we also observe that with higher values of m, the average score increases. This is because, with higher values of m, the reconstructed probability distribution is closer to the true distribution.

6.3 Time Cost

The time cost of making a prediction is also important. Table 4 shows the time cost for making a prediction using our PCTMC model with different parameters (we do not show the time costs for the Markov queueing model since they are negligible due to its small state space because of independence assumption). For real time application, we assume that the time cost of making a prediction must be less than one second. Thus, for point prediction, we recommend to set

Table 4. Time cost to make a prediction with 95 % confidence interval

	10 min	40 min	
PCTMC with $\theta = 0.03$	1.76 ± 0.2ms	6.98 ± 0.77ms	$m = 1$
	103 ± 13.7 ms	328 ± 43 ms	$m = 2$
	2.2 ± 0.2 s	8.9 ± 0.83 s	$m = 3$
PCTMC with $\theta = 0.02$	4.25 ± 0.4 ms	15.72 ± 1.42 ms	$m = 1$
	251 ± 25.5 ms	1.1 ± 0.1 s	$m = 2$
	8.9 ± 1.2 s	37 ± 3.5 s	$m = 3$
PCTMC with $\theta = 0.01$	13.5 ± 0.9 ms	49.1 ± 3.92 ms	$m = 1$
	8.8 ± 1.1 s	30.1 ± 0.31 s	$m = 2$
	33.9 ± 5.4 s	157 ± 17.8 s	$m = 3$

$\theta = 0.01, m = 1$ for both prediction ranges. For probability distribution prediction, we recommend to set $\theta = 0.02, m = 2$ for short range prediction, $\theta = 0.03, m = 2$ for long range prediction. Note that we used an Intel CORE i7 laptop with 8 GB RAM to run our experiments, the time cost could be considerably reduced if a more powerful machine, e.g. a server, were used.

7 Conclusion

We have presented a moment-based approach to make predictions of availability in bike-sharing systems. The moments of the number of available bikes are automatically derived via a PCTMC with time-inhomogeneous rates, fitted from historical data. The entire probability distribution is reconstructed using a maximum entropy approach. Our model is easy to understand since it explicitly captures the dynamics of the bike-sharing system. We demonstrated that it outperforms the classic Markov queueing model in several performance metrics for prediction accuracy. Moreover we have also shown that by using the direct contribution graph and the contribution propagation method, the model size can be significantly reduced to such an extent that it is suitable for real time application.

In future work we plan to explore the impact of neighbouring stations, and extend our model to capture their effects. For example, if a station is empty, then the user is likely to pick up a bike from a neighbouring station, thus increasing the pickup rate at the neighbouring station. Conversely, if a station is full, then the user is likely to return a bike to a neighbouring station, increasing the bike arrival rate there. We think another merit of our PCTMC model is that it can be easily extended to capture such impact by using the indicator function to check whether a neighbouring station is empty or full in order to alter the bike arrival and pickup rate of a station. Unfortunately we do not currently have data to capture the impact of neighbouring stations.

Acknowledgement. This work is supported by the EU project QUANTICOL, 600708.

References

1. Fishman, E.: Bikeshare: a review of recent literature. Transp. Rev. **36**(1), 1–22 (2015)
2. Lin, J.R., Yang, T.H.: Strategic design of public bicycle sharing systems with service level constraints. Transp. Res. Part E: Logist. Transp. Rev. **47**(2), 284–294 (2011)
3. Pfrommer, J., Warrington, J., Schildbach, G., Morari, M.: Dynamic vehicle redistribution and online price incentives in shared mobility systems. IEEE Trans. Intell. Transp. Syst. **15**(4), 1567–1578 (2014)
4. Nair, R., Miller-Hooks, E.: Fleet management for vehicle sharing operations. Transp. Sci. **45**(4), 524–540 (2011)
5. Contardo, C., Morency, C., Rousseau, L.M.: Balancing a Dynamic Public Bike-Sharing System, vol. 4. CIRRELT, Montreal (2012)
6. Schuijbroek, J., Hampshire, R., van Hoeve, W.J.: Inventory rebalancing and vehicle routing in bike sharing systems. In: Technical report, Schuijbroek (2013)
7. Yoon, J.W., Pinelli, F., Calabrese, F.: Cityride: a predictive bike sharing journey advisor. In: 2012 IEEE 13th International Conference on Mobile Data Management (MDM), pp. 306–311. IEEE (2012)
8. Gast, N., Massonnet, G., Reijsbergen, D., Tribastone, M.: Probabilistic forecasts of bike-sharing systems for journey planning. In: The 24th ACM International Conference on Information and Knowledge Management (CIKM 2015) (2015)
9. Froehlich, J., Neumann, J., Oliver, N.: Sensing and predicting the pulse of the city through shared bicycling. IJCAI **9**, 1420–1426 (2009)
10. Kaltenbrunner, A., Meza, R., Grivolla, J., Codina, J., Banchs, R.: Urban cycles and mobility patterns: exploring and predicting trends in a bicycle-based public transport system. Pervasive Mob. Comput. **6**(4), 455–466 (2010)
11. Guenther, M.C., Bradley, J.T.: Journey data based arrival forecasting for bicycle hire schemes. In: Dudin, A., De Turck, K. (eds.) ASMTA 2013. LNCS, vol. 7984, pp. 214–231. Springer, Heidelberg (2013)
12. Engblom, S.: Computing the moments of high dimensional solutions of the master equation. Appl. Math. Comput. **180**(2), 498–515 (2006)
13. Raviv, T., Kolka, O.: Optimal inventory management of a bike-sharing station. IIE Trans. **45**(10), 1077–1093 (2013)
14. Feng, C., Hillston, J., Galpin, V.: Automatic moment-closure approximation of spatially distributed collective adaptive systems. ACM Trans. Model. Comput. Simul. (TOMACS) **26**(4), 26 (2016)
15. Mead, L.R., Papanicolaou, N.: Maximum entropy in the problem of moments. J. Math. Phys. **25**(8), 2404–2417 (1984)
16. Andreychenko, A., Mikeev, L., Wolf, V.: Model reconstruction for moment-based stochastic chemical kinetics. ACM Trans. Model. Comput. Simul. (TOMACS) **25**(2), 12 (2015)
17. Tari, Á., Telek, M., Buchholz, P.: A unified approach to the moments based distribution estimation – unbounded support. In: Bravetti, M., Kloul, L., Zavattaro, G. (eds.) EPEW/WS-EM 2005. LNCS, vol. 3670, pp. 79–93. Springer, Heidelberg (2005)

18. Snyman, J.: Practical Mathematical Optimization: An Introduction to Basic Optimization Theory and Classical and New Gradient-Based Algorithms, vol. 97. Springer Science & Business Media, Heidelberg (2005)
19. Reinecke, P., Krauss, T., Wolter, K.: Hyperstar: phase-type fitting made easy. In: 2012 Ninth International Conference on Quantitative Evaluation of Systems, pp. 201–202. IEEE (2012)