Moment-based Availability Prediction for Bike-Sharing Systems

Jane Hillston

Joint work with Cheng Feng and Daniël Reijsbergen
LFCS, School of Informatics, University of Edinburgh
http://www.quanticol.eu

9th May 2018

Bike-sharing Systems (BSS)

- Bike-sharing system: users can rent bicycles for trips between stations.
- There are over 700 bike-sharing systems across the world.
- Biggest systems worldwide: Wuhan and Hangzhou, with 90,000 and 60,600 bicycles respectively.
- London Cycle Hire Scheme: over 11,000 bicycles and 750 stations.

Smart BSS

Data-driven modelling of BSS for

- Policy design: price, location of stations, etc.
- Intelligent bike redistribution: optimal route for bike redistribution.
- User journey planning: make predictions about trip feasibility in order to enhance user experience.

Our work focuses on user journey planning.

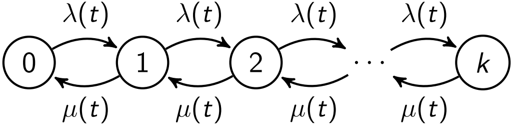
User Journey Planning

$X(t+h)$: a rv denotes the number of available bikes/slots at a station at a future time point $t+h$

- Literature focuses on point estimates: $\mathbb{E}[X(t+h)]$
- It is more informative to provide $\Pr(X(t+h) > N)$

Markov Queueing Model

Idea: split a day into n slots, fix the bike arrival and pickup rates of a station in a single slot, then a station can be modelled as a time-inhomogeneous Markov queue $M/M/1/k$



$\lambda(t)$, $\mu(t)$: the time-dependent bike arrival and pickup rates of the station at time t, can be learned from historical data.

Let $Q(t)$ be the generator matrix of the CTMC at time t

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

Markov Queueing Model cont.

The Markov queueing model assumes the state of a particular station does not depend on the state of the others, thus stations can be modelled in isolation.

This 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.

A more realistic model should also capture the journey dynamics between stations.

Hence, we propose a time-inhomogeneous Population CTMC model, which captures the journey dynamics between stations.

Time-inhomogeneous PCTMC

A stochastic process which can be expressed as a tuple $\mathcal{P} = (\mathbf{X}, \mathcal{T}, \mathbf{x_0})$:

- $\mathbf{X} = (x_1, ..., x_n) \in \mathbb{Z}^n_{\geq 0}$ is an integer vector with the ith $(1 \leq i \leq n)$ component representing the current population of an agent type S_i.
- $\mathcal{T} = \{\tau_1, ..., \tau_m\}$ is the set of transitions, 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 the time-dependent rate function.
  2. $\mathbf{d}_\tau \in \mathbb{Z}^n$ is the update vector.
- $\mathbf{x_0} \in \mathbb{Z}^n_{\geq 0}$ is the initial state of the model.

Moment approximation for PCTMCs

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

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

where the net change on the population of agent type S_i due to transition τ is given by $d_\tau^i = \overline{\ell}_i - \underline{\ell}_i$ $(1 \leq i \leq n)$.

The evolution of population moments of an arbitrary PCTMC model can be approximated by the following system of ODEs

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

where $M(\mathbf{X})$ denotes the moment to be calculated.

Moment Equations for PCTMCs

Replacing $M(\mathbf{X})$ with x_i, x_i^2, $x_i x_j$, we get the moment equations for the first moment, second moment and second-order joint moment:

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[x_i] &= \sum_{\tau \in \mathcal{T}} d_\tau^i \mathbb{E}[r_\tau] \\
\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[{x_i}^2] &= 2\sum_{\tau \in \mathcal{T}} d_\tau^i \mathbb{E}[x_i \times r_\tau] + \sum_{\tau \in \mathcal{T}} {d_\tau^i}^2 \mathbb{E}[r_\tau] \\
\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[x_i x_j] &= \sum_{\tau \in \mathcal{T}} d_\tau^i \mathbb{E}[x_j \times r_\tau] + \sum_{\tau \in \mathcal{T}} d_\tau^j \mathbb{E}[x_i \times r_\tau] + \sum_{\tau \in \mathcal{T}} d_\tau^i \times d_\tau^j \mathbb{E}[r_\tau]
\end{aligned}
$$

The system of ODEs can be solved rather efficiently by numerical simulation if the system is closed, otherwise moment-closure techniques need to be applied to close the system before solving the ODEs.

Model BSS as PCTMC

NB: Journey durations are fitted by Erlang distributions.

A naive PCTMC model for a BSS consisting of N stations:

$$\forall i, j \in (1, N)$$

$$Bike_i \longrightarrow Slot_i + Journey_j^i@P_1 \quad \text{at } \mu_i(t)p_j^i(t)$$

$$Journey_j^i@P_l \longrightarrow Journey_j^i@P_{l+1} \quad \text{at } (P_j^i/d_j^i)\,\#(Journey_j^i@P_l)$$

$$l \geq 1 \wedge l < P_j^i$$

$$Journey_j^i@P_{P_j^i} + Slot_j \longrightarrow Bike_j \quad \text{at } (P_j^i/d_j^i)\,\#(Journey_j^i@P_{P_j^i})$$

$Journey_j^i@P_l$: a bike agent which is currently on a journey from station i to station j at phase l, $1 \leq l \leq P_j^i$.

The Naive PCTMC model for BSS

Problems

$$\forall i, j \in (1, N)$$

$$Bike_i \longrightarrow Slot_i + Journey_j^i @ P_1 \quad \text{at } \mu_i(t) p_j^i(t)$$

$$Journey_j^i @ P_l \longrightarrow Journey_j^i @ P_{l+1} \quad \text{at } (P_j^i / d_j^i) \, \#(Journey_j^i @ P_l)$$

$$l \geq 1 \wedge l < P_j^i$$

$$Journey_j^i @ P_{P_j^i} + Slot_j \longrightarrow Bike_j \quad \text{at } (P_j^i / d_j^i) \, \#(Journey_j^i @ P_{P_j^i})$$

N is usually very large, derived system of ODEs are infeasible to solve.

Only make prediction for a single target station at a time.

Prune the PCTMC to a reduced one in which only stations with significant journey flows to the target station are modelled explicitly.

Prune PCTMC for BSS

general idea

- Use contribution coefficient C_{ij} to quantify the contribution of station j to the journey flows to station i.
- Regard station j as significant with respect to station i if the contribution coefficient C_{ij} is above a specific threshold.
- Only model those significant stations explicitly in the reduced PCTMC.

Direct contribution coefficient

Contribution on journey flows of one station to another can be both direct and indirect.

The definition of a direct contribution coefficient at time t is given by the following simple formula:

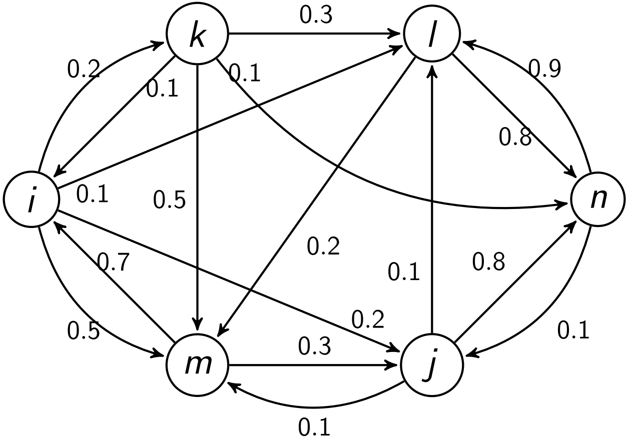
$$c_{ij}(t) = \lambda_i^j(t)/\lambda_i(t)$$

$\lambda_i^j(t)$: the bike arrival rate from station j to station i at time t.

$\lambda_i(t) = \sum_j \lambda_i^j(t)$.

Directed Contribution Graph

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 $c_{ij} > 0$.



Indirect contribution coefficient

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:

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

The contribution coefficient of station j to station i is characterized by the maximum of the path dependent coefficients:

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

Derive significant station set

Given a target station v, then for $i \in (1, 2, \ldots, N)$, we can infer:

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

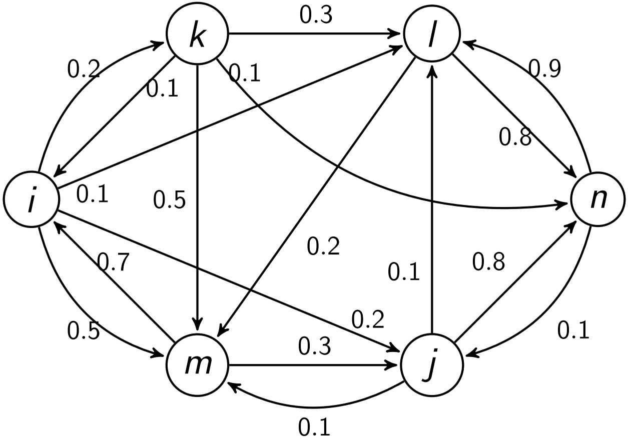
$\Theta(v)$: the set of bike stations in which all stations have a significant contribution to the journey flows to a given target station v for bike availability prediction.

$\theta \in (0, 1)$: a threshold value which can be used to control the extent of model reduction. On average, more than 96% stations can be excluded if θ is set to value 0.01 for London BSS.

Derive significant stations set

cont.

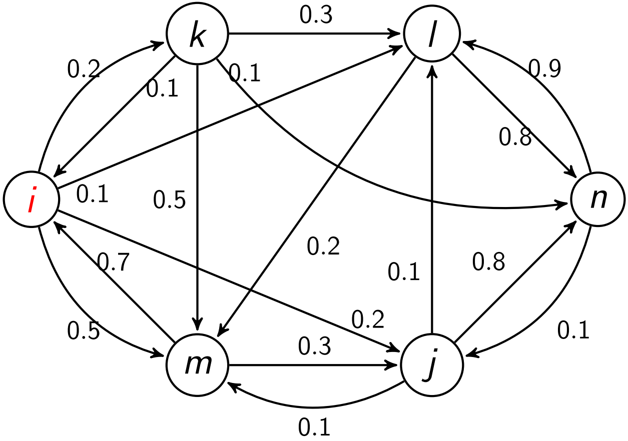
Suppose $\theta = 0.2$



Derive significant stations set

cont.

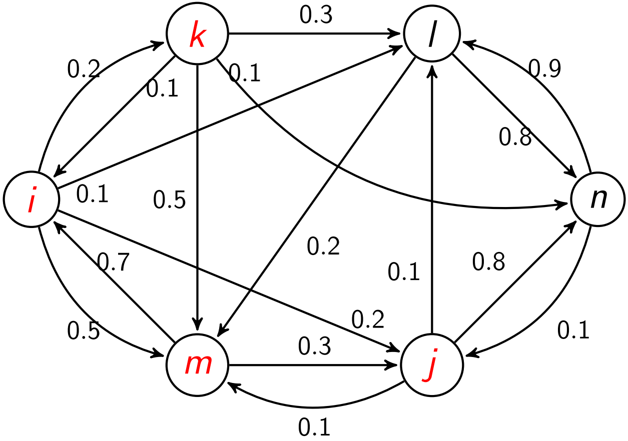
Suppose $\theta = 0.2$



Derive significant stations set

cont.

Suppose $\theta = 0.2$



Derive significant stations set

cont.

- Input: a target station v, current time t and prediction horizon h.
- Output: $\Theta(v) = \Theta(v, s_1) \cup \Theta(v, s_2) \cup \ldots \cup \Theta(v, s_n) \cup v$.

$(s_1, s_2, \ldots, s_n)$: the minimal set of time slots which covers $[t, t + h]$.

$\Theta(v, s_i)$: the set of significant stations for the target station in time slot s_i.

The reduced PCTMC for BSS

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

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

The reduced PCTMC for BSS

cont.

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

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

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

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

Specify the initial state of the reduced PCTMC

Given a snapshot of the bike-sharing system at a time instant t which contains the following information:

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

Let

$$Journey^i(t, \Delta t) = Journey^i_j(t, \Delta t)$$

if

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

α is a random number uniformly distributed in $(0, 1)$.

Specify the initial state of the reduced PCTMC

Given a snapshot of the bike-sharing system at a time instant t which contains the following information:

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

Let

$$Journey_j^i(t, \Delta t) = Journey_j^i @ P_l$$

if

$$\Delta t \geq (l-1)d_j^i / P_j^i \text{ and } \Delta t < l \times d_j^i / P_j^i,$$

where $l \leq P_j^i$. Otherwise, if $l > P_j^i$, we let

$$Journey_j^i(t, \Delta t) = Journey_j^i @ P_{P_j^i}$$

.

From Moments to Probability Distribution

By solving the system of moment ODEs of the reduced PCTMC for BSS, we obtain the first m moments of the number of available bikes X_v in the target station: $(u^1, u^2, \ldots, u^m)$

Our goal is to is to reveal the full probability distribution of X_v.

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.

Let $\mathcal{G}$ be the set of all possible probability distributions for X_v, we select a distribution g which maximizes the entropy $H(g)$ over all distributions in $\mathcal{G}$:

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

s.t.

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

where $u^0 = 1$

This is a constrained optimization problem.

Probability Distribution Reconstruction

Maximum Entropy Approach cont.

Introduce one Lagrange multiplier λ_n per moment constraint, we seek the 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 $g(x)$ yields:

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

Substitute $g(x)$ back into the Lagrangian, the problem is transformed into an unconstrained minimization problem with respect to variables $\lambda_1, \lambda_2, \ldots, \lambda_n$:

$$\arg\min \Gamma(\lambda_1, \lambda_2, \ldots, \lambda_n) = \ln \sum_{x=0}^{k_v} \exp\left(-\sum_{n=1}^{m} \lambda_n x^n \right) + \sum_{n=1}^{m} \lambda_n u^n$$

$\Gamma(\lambda_1, \lambda_2, \ldots, \lambda_n)$ is a convex function, thus there exists a unique solution to minimize Γ.

No analytical solution, but a close approximation $(\lambda_1^*, \lambda_2^*, \ldots, \lambda_n^*)$ can be found through gradient descent, and we can finally predict

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

Experiments

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.

For parameter estimation, we split a day into slots of 20 minute duration.

Prediction Horizon is set to 10 minutes for short range prediction and 40 minutes for long range prediction.

Given a vector **x** of predictions and **y** of observations, with A the set of prediction/observation pairs, the RMSE is defined as

$$\sqrt{\frac{1}{|A|}\sum_{i \in A}(x_i - y_i)^2}$$

The calculated RMSE on the prediction of the number of available bikes:

	10min	40min	
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	$m = 1, 2, 3$

- RMSE works for point predictions.
- Not suitable for evaluating trip feasibility predictions.

A proper score rule proposed by Gast *et al.* to evaluate trip feasibility predictions:

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

Includes a penalty of 4 for incorrectly recommending to go when there is no bike available, a penalty of $\frac{1}{4}$ for incorrectly recommending not to go when there is a bike available

Evaluation

Average score of making a recommendation to "Will there be a bike?" query

	10min	40min	
Markov queueing model	0.90 ± 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	$m = 3$

Evaluation

Average score of making a recommendation to the "Will there be a slot?" query

	10min	40min	
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	$m = 3$

Evaluation

Time cost to make a prediction

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

Conclusion and future work

- Using a moment-based approach to make prediction of bike availability capturing significant journey dynamics between stations can achieve better accuracy, compared with Markov queueing model which analyses stations in isolation.
- The moment-based approach is suitable for real time application.
- Future work: explore the impact of neighbouring stations, and extend our model to capture their effects.