Omega 65 (2016) 10–16

Contents lists available at ScienceDirect

Omega

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

Detection of unusable bicycles in bike-sharing systems☆

Mor Kaspi, Tal Raviv*, Michal Tzur

Department of Industrial Engineering, Faculty of Engineering, Tel-Aviv University, Tel-Aviv 69978, Israel

ARTICLE INFO

Article history:
Received 21 July 2015
Accepted 8 December 2015
Available online 17 December 2015

Keywords:
Bike-sharing
Maintenance
Bayesian model

ABSTRACT

In bike-sharing systems, a small percentage of the bicycles become unusable every day. Currently, there is no reliable on-line information that indicates the usability of bicycles. We present a model that estimates the probability that a specific bicycle is unusable as well as the number of unusable bicycles in a station, based on available trip transaction data. Further on, we present some information based enhancements of the model and discuss an equivalent model for detecting locker failures.

1. Introduction

Bike-sharing systems have become a common sight in many cities around the world during the last decade. In some of the major cities, this mode of transportation attracts a considerable amount of commuters and tourists on a daily basis. For example, the bike-sharing system in New-York, CitiBike, reported on an average of 34,176 rides per day during August 2014 [8].

Bike-sharing systems are typically subsidized and regulated by the local governments. Such systems should be designed and operated in the most efficient possible way. The two main components of the operating costs are due to repositioning and maintenance activities. The planning of the repositioning activities has received substantial attention in the literature, see, for example [2,3] and the references therein.

Maintenance operations of bike-sharing systems have not been so far at the focus of Operation Research or Operations Management studies. We envision a framework for the planning of these operations that includes three processes: (1) detection of unusable bicycles; (2) analysis of the effect of the presence of unusable bicycles on the quality of service provided to the users; (3) collection of unusable bicycles to maintenance shops or repairing them on-site. The first process is at the focus of this note, while the following two are studied in [5,6], respectively.

The information systems installed in bike-sharing systems present to the public on-line aggregated information about each station. In particular, using smartphones or stations' kiosks, users may query the state of each station in terms of the number of available bicycles and the number of available lockers. Internally, the information system stores a log of the transactions that were carried out. Each transaction is featured by its type (renting, repositioning, maintenance), start time, end time, start station ID, start locker ID, end station ID, end locker ID, bike ID, User ID (either a regular user of the system or a maintenance personnel), etc. Some operators share a subset of the fields in this log with the public, see for example the CitiBike trip history: http://www.citibikenyc.com/system-data.

In existing bike-sharing systems, information about unusable bicycles is received either from users or from of repositioning workers when they service the stations. The probability that a user will report on an unusable bicycle is rather low if other bicycles that are parked in the station can be rented. That is, a user will typically complain about an unusable bicycle when there is no alternative in the station. In addition, not all stations are serviced by repositioning workers on a daily basis. Therefore an unusable bicycle may be parked at a station for a long period of time before being detected and collected.

In some systems, such as CitiBike, each locker is equipped with a maintenance button that the users may push in order to signal to the operator that the bicycle should be serviced. While through this mechanism, more information about bicycles that should be repaired is obtained, this also generates a fair amount of false alarms. In the CitiBike system, about 36% of the reported bicycles are actually usable [7] and, more importantly, many unusable bicycles are not reported through this button by the users.

Undetected unusable bicycles appear in the information systems as available ones. This inaccuracy may adversely affect user's route choices and result in an inferior service level. For example, a user may go to a station with such undetected unusable bicycles only to find out that there are actually no available usable bicycles in the station. If the system could provide her with accurate information in advance she could save time by planning her trip

☆This manuscript was processed by Associate Editor Pesch.

* Corresponding author. Tel.: +972 3 6406977.
E-mail addresses: morkaspi@post.tau.ac.il (M. Kaspi), talraviv@eng.tau.ac.il (T. Raviv), tzur@eng.tau.ac.il (M. Tzur).

http://dx.doi.org/10.1016/j.omega.2015.12.003
0305-0483/

differently, e.g., start her trip at a neighboring station or select a different mode of transportation.

The contribution of this note is as follows: we propose using data that is already collected by existing bike sharing systems to estimate the probability that each bicycle is usable. We formulate a Bayesian model that makes use of on-line transactions data to constantly update these probabilities and propose a method to approximate these probabilities in real-time. Subsequently, we present some possible extensions of the model and explain how additional information such as user complaints can be incorporated in the model. In addition, we discuss how an equivalent model can be used for detection of locker (dock) failures.

2. A Bayesian model

The goal of this study is to estimate the number of unusable bicycles in a station and to continuously update this estimation in real-time. We begin by focusing on each bicycle independently. We assign a *Probability of Unusability* (PoU) to each bicycle in the system and update it continuously. A good indication for unusability of a bicycle is the fact that it was not rented for a long period. However, this probability also depends on other factors such the number of renting transactions since the bicycle arrived at the station and the availability of other bicycles in the station when these transactions occurred. The model that will be presented next makes use of the transaction data in order to estimate the PoU of each bicycle in a single station.

We use the following notation:

i	Bicycle ID
e	Rent event
C	Set of all lockers in the station, $\lvert C\rvert$ is the capacity of the station
p_i	Prior probability that bicycle i is returned to the station unusable
S^e	Set of bicycles that are parked in the station right before rent event e
$q^e(m,S)$	The probability that right before rent event e there are m usable bicycles in the set S
$P^e(x)$	The probability of scenario x at rent event e
$P^e(x,y)$	The joint probability of scenarios x and y at rent event e
p_i^e	The PoU of bicycle i right after the occurrence of rent event e

We assume that when a bicycle is rented, it is usable, that is, a user never rents an unusable bicycle. Formally, we assume $P^e(i\ usable \mid i\ rented)=1$ and $P^e(i\ rented \mid i\ unusable)=0$. However, bicycle i may turn unusable during a ride, and therefore there is a probability p_i that the bicycle will be returned to the station unusable. See discussion in Section 5.3 regarding the calculation of this probability.

For simplicity of the presentation, we initially assume that the users have no preferences regarding the locker from which the bicycle will be rented. That is, a user uniformly selects a bicycle from the usable bicycles that are parked in the station. In Section 5.1, we discuss how user preferences regarding the lockers can be incorporated in the model.

Our goal is to update the PoU of bicycles that are parked in the station. Given that at rent event e bicycle j was rented, we use Bayes' rule to calculate the probability that bicycle i ($i \neq j$) is unusable. This calculation is carried out for any bicycle that is left parked at the station:

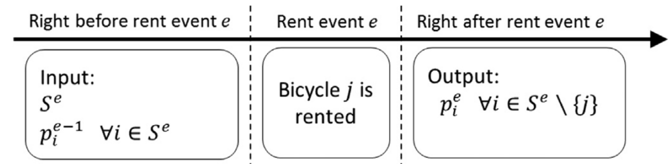

Fig. 1. Updating the PoU at rent event e.

$$p_i^e = P^e(i\ unusable \mid j\ rented) = \frac{P^e(i\ unusable, j\ rented)}{P^e(j\ rented)} \quad \forall i \in S^e \setminus \{j\} \tag{1}$$

Fig. 1 depicts the notation used right before, at and right after rent event e. The updating of the PoU is carried out for bicycles that are left parked in the station right after each rent.

To calculate (1), let us consider first the denominator. Given that bicycle j is usable, the probability that it will be rented at rent event e is given by:

$$P^e(j\ rented \mid j\ usable) = \sum_{m=0}^{|S^e|-1} \frac{1}{1+m} \cdot q^e\left(m, S^e \setminus \{j\}\right) \tag{2}$$

This expression is obtained by conditioning on the number of usable bicycles in the station (m), excluding bicycle j, and multiplying the probability of having this number, $q^e(m, S^e \setminus \{j\})$, by the uniform probability that j will be selected from within $m+1$ usable bicycles.

Then, by definition: $P^e(j\ usable \mid j\ rented)=1$, and equivalently: $P^e(j\ rented, j\ usable) = P^e(j\ rented)$. Using Bayes' rule, we obtain the probability that bicycle j will be rented at event e:

$$P^e(j\ rented) = P^e(j\ rented, j\ usable) = P^e(j\ usable) \cdot P^e(j\ rented \mid j\ usable)$$

$$= \left(1 - p_j^{e-1}\right) \cdot \sum_{m=0}^{|S^e|-1} \frac{1}{1+m} \cdot q^e\left(m, S^e \setminus \{j\}\right) \tag{3}$$

where p_j^{e-1} denotes the PoU of bicycle j right before rent event e. Similarly to calculate nominator of (1) we condition in addition over bicycle i and obtain the joint probability that bicycle i is unusable and bicycle j is rented at rent event e:

$$P^e(i\ unusable, j\ rented) = p_i^{e-1} \cdot \left(1 - p_j^{e-1}\right) \cdot \sum_{m=0}^{|S^e|-2} \frac{1}{1+m} \cdot q^e\left(m, S^e \setminus \{i,j\}\right) \tag{4}$$

Eqs. (3) and (4) contain an expression for the probability of the number of usable bicycles m, excluding i and j, that are parked in the station right before rent event e. Note that each bicycle that is parked in the station has a different probability of being usable. Therefore, this expression is the sum of Bernoulli variables with different success probabilities, which is a Poisson Binomial distribution (see, for example, [4]), given by the following probability mass function:

$$q^e\left(m, S^e\right) = \sum_{S \in F_m(S^e)} \prod_{k \in S} \left(1 - p_k^{e-1}\right) \prod_{k \in S^e \setminus S} p_k^{e-1}$$

where $F_m(S)$ denotes the collection of all subsets of set S with cardinality m. Note that in Eqs. (3) and (4) this probability is calculated for all possible values of m and therefore the calculation effort for evaluating (3) and (4) grows exponentially in S^e. Thus, the exact on-line updating of the PoU is impractical for large bike sharing stations. However, we observe that a related quantity, the expected number of usable bicycles, is easier to calculate as it is merely the sum of the probabilities of usability of the bicycles in the station:

$$\mathbb{E}\left(usables\ in\ S^e\right) = \sum_{i \in S^e} \left(1 - p_i^{e-1}\right)$$

In the following section we propose a method to approximate the PoU, based on the expected number of usable bicycles in the station.

3. Approximating the probability of unusability

Henceforth, we denote the approximated PoU of bicycle i after rent event e by $\tilde{p}_i^e$. Recall that we assume that bicycles are selected uniformly from the set of available usable bicycles in the station. We approximate the probability that bicycle j is rented, given that it is usable, by assuming that the number of usable bicycles in the station is known and equals its expectation. Specifically, given that bicycle j is usable, the expected number of usable bicycles in the station is one plus the expected number of bicycles in the remaining set of available bicycles. Thus, Eq. (2) is approximated as follows:

$$\tilde{P}^e(j\ rented|j\ usable) = \frac{1}{\tilde{\mathbb{E}}(usables\ in\ S^e|\ j\ usable)} = \frac{1}{1+\sum_{k\in S^e\setminus\{j\}}\left(1-\tilde{p}_k^{e-1}\right)} \tag{5}$$

where the expected number of usable bicycles in the station right before rent event e is approximated by:

$$\tilde{\mathbb{E}}(usables\ \text{in}\ S^e \setminus \{j\}) = \sum_{k\in S^e\setminus\{j\}}\left(1-\tilde{p}_k^{e-1}\right) \tag{6}$$

Similarly to the calculation of Eq. (3), we multiply Eq. (5) by the probability that bicycle j is usable, to obtain the approximated probability that bicycle j is rented at rent event e:

$$\tilde{P}^e(j\ rented) = \tilde{P}^e(j\ rented, j\ usable) = \tilde{P}^e(j\ usable)\cdot\tilde{P}^e(j\ rented|j\ usable)$$
$$= \frac{1-\tilde{p}_j^{e-1}}{1+\sum_{k\in S^e\setminus\{j\}}\left(1-\tilde{p}_k^{e-1}\right)} \tag{7}$$

Next, given that bicycle i is unusable, the expected number of usable bicycles in the station is updated to exclude i and thus we obtain an approximation of the following conditional probability:

$$\tilde{P}^e(j\ rented|i\ unusable) = \frac{1-\tilde{p}_j^{e-1}}{1+\sum_{k\in S^e\setminus\{i,j\}}\left(1-\tilde{p}_k^{e-1}\right)}$$

And again, using Bayes' rule, we obtain an approximation of the joint probability:

$$\tilde{P}^e(j\ rented, i\ unusable) = \frac{\left(1-\tilde{p}_j^{e-1}\right)\cdot\tilde{p}_i^{e-1}}{1+\sum_{k\in S^e\setminus\{i,j\}}\left(1-\tilde{p}_k^{e-1}\right)} \tag{8}$$

Finally, by dividing Eq. (8) by Eq. (7) we obtain an approximation of the updated PoU:

$$\tilde{p}_i^e = \tilde{P}^e(i\ unusable|j\ rented) = \tilde{p}_i^{e-1}\cdot\frac{1+\sum_{k\in S^e\setminus\{j\}}\left(1-\tilde{p}_k^{e-1}\right)}{1+\sum_{k\in S^e\setminus\{i,j\}}\left(1-\tilde{p}_k^{e-1}\right)} \tag{9}$$

Using Eq. (6), we can rewrite Eq. (9) as:

$$\tilde{p}_i^e \equiv \tilde{P}^e(i\ unusable|j\ rented) = \tilde{p}_i^{e-1}\cdot\frac{1+\tilde{\mathbb{E}}(usables\ in\ S^e\setminus\{j\})}{\tilde{p}_i^{e-1}+\tilde{\mathbb{E}}(usables\ in\ S^e\setminus\{j\})}$$

We observe that the PoU of a bicycle increases after every rent event in which it is not selected. In addition, as the initial PoU of a bicycle is its prior probability, it is easy to see that the PoU of a bicycle increases also as its prior probability increases.

So far, our focus was on calculating the PoU for each bicycle separately. However, it is typically more interesting for the operators and the users to view all the bicycles in a station aggregately. Similar to the above, our analysis also provides the expected number of unusable bicycles in the station, given in by:

$$\tilde{\mathbb{E}}(unusables\ in\ S^e) = \sum_{i\in S^e}\tilde{p}_i^{e-1} \tag{10}$$

We note that the value calculated in (10) can be used as a reliable estimator for the actual number of unusable bicycles after a sufficient number of rent events. In the next section this will be demonstrated numerically. The expected number of unusable bicycles is a more concise measure, compared to the PoU of each bicycle in the station. It can be used and understood by the operators and by the users.

The approximation of the PoU and the expected number of unusable bicycles in a station can be carried out after each rent event in $O(|C|)$ time, i.e., linearly in the station capacity. At each rent event, these values can be updated in a fraction of a second. Therefore, the estimated number of unusable bicycles can be displayed on-line to the operators and the users. In the following section, we show that this is a very accurate approximation by comparing it to the result of the exact calculation for small stations with up to 15 lockers. Note that the complexity of the exact method is $O\left(|C|\cdot 2^{|C|}\right)$ for each rent event, which is impractical for on-line usage.

4. Numerical results

In this section, we present the results of a numerical experiment carried out to test our proposed detection model. To simulate the on-line calculation of the PoU, we have used CitiBike trip history transactions data from July–August 2014. Using this data, we estimated the renting/returning rates on weekdays in each station during 30 min periods along the day. We generated 100 demand realizations per station. Each demand realization consists of a set of renting and returning events and their times of occurrence along a 2-day period. Each return event is supplemented with a binary parameter that indicates whether the bicycle is usable or not. For the experiment, we set the failure probability of all bicycles to 0.01. That is, the unusability indicator value was drawn from a Bernoulli distribution with parameter 0.01. In addition, we assume that at the initial state of the station all bicycles parked at the station are usable, as if replenishment activities and collection of unusables were just executed. The initial inventory level is set to the optimal level according to the method of Kaspi et al. [5]. The demand realizations data used in the simulation can be downloaded from http://www.eng.tau.ac.il/~morkaspi/publications.html.

At a rent event, if there are available usable bicycles in the station, one is selected uniformly and is removed from the set of available bicycles. If there are no available usable bicycles in the station the user is assumed to abandon the station. At return events, if there are available lockers in the station, one is selected uniformly and the bicycle is returned to that locker. If there are no available lockers in the station the user is assumed to abandon the station. If an unusable bicycle is returned to the station, it "occupies" a locker, but is not entered to the set of available usable bicycles in a station (and therefore will never be selected at a rent event).

We compare the approximated expected number of unusable bicycles (Eq. (10)) to a naïve approach for assessing the expected number of unusable bicycles in a station. The naïve expectation is obtained by summing the prior probabilities of the bicycles returned to a station in a given time period. For example, assume that the prior probability of all bicycles is 0.01 and that in a given time period 200 bicycles were returned to the station. The naïve estimation of the number of unusable bicycles that were returned

Table 1
Simulation results – average over 100 realizations for 20 stations in CitiBike.

Station ID	Capacity	Number of rents	Number of returns	Unusable bicycles	Naïve MAD (Stdev)	PoU MAD (Stdev)	PoU better (out of 100)	*P*-value
134	35	393.67	403.61	4.34	1.56 (1.92)	0.39 (0.61)	85	< 0.0001
145	36	392.96	404.82	4.09	1.38 (1.67)	0.50 (0.67)	80	< 0.0001
132	35	366.35	364.71	3.85	1.42 (1.77)	0.37 (0.58)	87	< 0.0001
135	42	251.89	256.65	2.54	1.22 (1.50)	0.14 (0.30)	94	< 0.0001
139	35	208.78	206.55	1.90	1.13 (1.38)	0.15 (0.31)	88	< 0.0001
144	27	179.65	177.77	1.52	0.98 (1.20)	0.31 (0.48)	92	< 0.0001
142	42	176.33	182.24	1.89	1.11 (1.40)	0.32 (0.51)	87	< 0.0001
133	27	163.04	175.27	1.93	1.04 (1.31)	0.54 (0.66)	73	< 0.0001
138	39	108.53	110.37	1.30	0.84 (1.11)	0.21 (0.42)	80	< 0.0001
137	28	100.81	112.66	1.13	0.84 (1.06)	0.60 (0.74)	59	0.0284
136	23	82.74	94.71	0.87	0.62 (0.84)	0.44 (0.51)	54	0.1841
127	27	76.70	90.85	0.89	0.69 (0.94)	0.46 (0.61)	57	0.0666
128	19	67.45	77.97	0.78	0.69 (0.86)	0.41 (0.50)	72	< 0.0001
129	29	65.71	82.16	0.74	0.66 (0.81)	0.45 (0.56)	62	< 0.0001
143	23	56.26	69.22	0.68	0.69 (0.81)	0.40 (0.53)	78	< 0.0001
140	27	38.30	45.36	0.43	0.56 (0.67)	0.38 (0.54)	71	< 0.0001
130	23	33.33	36.09	0.31	0.47 (0.54)	0.45 (0.53)	53	0.2421
126	24	32.51	33.60	0.31	0.46 (0.51)	0.31 (0.43)	83	< 0.0001
141	31	31.81	40.88	0.48	0.56 (0.69)	0.46 (0.61)	73	< 0.0001
131	23	24.82	34.03	0.35	0.48 (0.55)	0.34 (0.48)	82	< 0.0001

to the station would be 2. We note that this unbiased estimator of the expected number of unusable bicycles by itself may provide a relatively good picture regarding the amount of unusable bicycles in a station. Given that in some bike sharing systems the number of unusable bicycles is not assessed at all, using even this naïve method would be valuable.

In Table 1 we present simulation results for 20 arbitrarily selected stations. Simulation results of another 80 stations are available online, as an electronic supplementary, at http://www.eng.tau.ac.il/~morkaspi/publications.html. In the first and second columns of Table 1, the station ID and capacity are presented, respectively. The average (over 100 realizations) of the number of bicycles that were rented and returned to the station in the simulation are presented in the third and fourth column. Note that the realized number of rent/return events was in most cases a bit larger, but not all bicycles could be rented/returned due to bicycle/locker shortages. In the fifth column, the average number of unusable bicycles that were returned to the station is presented. For each demand realization, we calculate at the end of the 2-days period the difference between the actual number of unusable bicycles and its estimation obtained by the naïve approach and by the PoU approach. The mean absolute deviation (and the standard deviation) of these differences are presented in the sixth and seventh columns, respectively. In addition, we count the number of times in which the PoU estimation was closer to the actual value as compared to the estimation of the naïve approach. This number is presented in the eighth column. The *P*-value of the sign-test used to determine whether the PoU approach generates closer estimation as compared to the naïve approach is presented in the last column.

As can be observed, both the mean absolute deviation and the standard deviation of the PoU approach are significantly smaller than those of the naïve approach. In particular, as more rent events occur in the station (the table is sorted in decreasing order of the number of rent events), more information is accumulated by the PoU approach and can be used to better estimate the number unusable bicycles. The figures in the eighth and last columns demonstrate the superiority of the PoU approach as compared to the naïve approach.

Recall that in the simulation, we set the actual failure probability to 0.01. The estimations presented in Table 1 are based on the assumption that indeed the prior probability is 0.01. In reality, the operator may not have an exact knowledge of the prior probabilities. Next, we examine whether the PoU approach results with better estimations as compared to the naïve approach even if the exact prior probabilities are unknown exactly. We conducted the following analysis: we used the same demand realizations as in Table 1 (using a failure probability of 0.01) but assumed different levels of prior probabilities in the calculation of the PoU and naïve based estimations.

Table 2
Sensitivity analysis – actual failure probability 0.01.

Station ID	Capacity	PoU better (out of 100)				
		Assumed prior				
		0.001	0.005	0.01	0.02	0.05
134	35	98	85	85	96	100
145	36	99	87	80	99	100
132	35	100	92	87	97	100
135	42	100	98	94	98	100
139	35	98	86	88	96	100
144	27	94	83	92	93	100
142	42	93	79	87	92	100
133	27	90	73	73	90	100
138	39	95	95	80	86	100
137	28	76	75	59	80	100
136	23	81	81	54	86	99
127	27	77	75	57	80	99
128	19	80	80	72	82	98
129	29	78	79	62	77	99
143	23	87	87	78	68	98
140	27	71	71	71	63	92
130	23	54	53	53	51	45
126	24	82	82	83	82	89
141	31	73	73	73	63	79
131	23	80	81	82	81	84

In Table 2, we present the number of times (out of 100 realizations) that the PoU based estimation was closer to the actual values as compared to the naïve approach. The first and second columns of Table 2 are identical to those of Table 1. In the third to seventh columns, we present these values for the following assumed prior probabilities: 0.001, 0.005, 0.01, 0.02, and 0.05, respectively.

Noticeably, as the assumed prior increases (or decreases) relative to the actual prior, the number of times the PoU approach delivers better estimations increases. This demonstrates that the

PoU approach is more robust with respect to the estimation of the prior probability as compared to the naïve approach and suggests that the model is not sensitive to the exact nature of the failure process.

A further validation of the detection model may be accomplished using real-time information regarding the actual number of unusable bicycles in certain points of time, and comparing it to the estimated one. Such information may be collected, for example, when repositioning or maintenance staff visit the station. Currently, the data required for such validation is not at our disposal. Moreover, it is not available to the system operators that we have been in contact with (CitiBike and Tel-O-Fun). To the best of our knowledge, this kind of data is not collected by any bike-sharing operator. Nevertheless, we emphasize that the demand data, stations' capacities and prior failure probabilities used in our simulation model were obtained from a real-world system (CitiBike). We believe that the characteristics of the simulation model are close enough to the behavior of the system in reality in order to demonstrate the effectiveness of the detection model.

Next, we present a numerical study carried out in order to test the accuracy of the approximated PoU (Section 3) as compared to the exact calculation (Section 2). We have used trip history transactions from the Capital Bikeshare system in Washington DC, during the 2nd quarter of 2013. We have selected 20 stations with capacities smaller than 15, for which the exact calculation could be done in a reasonable time. The preprocessing of the trip history data and the simulation were executed in the same manner as described above.

In Table 3 we present the simulation results. The first five columns present information about the realizations, as in Table 1. In the two rightmost columns, we present the average and the maximum absolute difference, over 100 realizations, between the exact and the approximated calculations of the expected number of unusable bicycles. Observing the rightmost two columns in the table, one can notice that the exact and the approximated calculation of the PoU result with very similar estimations of the number of unusable bicycles.

Table 3
Simulation results for 20 small stations in Capital Bikeshare.

Station ID	Capacity	Number of rents	Number of returns	Unusable bicycles	Average difference	Maximal difference
73	14	151.89	151.11	1.46	0.0079	0.1683
118	14	134.81	136.56	1.35	0.0059	0.1147
183	13	128.74	127.95	1.34	0.0057	0.0576
187	14	119.29	118.01	0.97	0.0066	0.1049
186	14	118.32	123.33	1.30	0.0085	0.0958
201	11	115.52	113.97	1.16	0.0120	0.1117
188	10	113.77	114.15	1.08	0.0081	0.0949
126	13	84.08	84.31	0.77	0.0056	0.0545
136	14	76.64	80.10	0.77	0.0098	0.1626
155	11	47.31	41.89	0.53	0.0113	0.1346
209	14	46.63	46.19	0.56	0.0077	0.1269
163	11	44.15	41.15	0.32	0.0092	0.0973
107	11	43.43	41.18	0.48	0.0108	0.1934
191	11	40.85	43.93	0.43	0.0079	0.0892
80	14	38.35	31.61	0.33	0.0091	0.1023
88	10	35.54	36.58	0.36	0.0166	0.1731
175	11	34.88	34.26	0.27	0.0065	0.1235
10	11	34.55	37.84	0.45	0.0148	0.1498
154	14	32.18	22.61	0.17	0.0030	0.0881
140	14	30.54	30.79	0.37	0.0140	0.1271

5. Extensions

In previous sections, we have made some simplifying assumptions regarding the available data and the user preferences, mainly for ease of the presentation. However, additional available information can be used to fine-tune the estimation of the PoU of each bicycle in the system. We now discuss some enhancements of the model.

5.1. User preferences

So far, we have assumed for simplicity that a renter selects uniformly a bicycle from within the set of usable bicycles in the station. However, in some stations, we observe that some lockers are much busier than others, probably due to their distance from the station's kiosk or due to their accessibility to pedestrians. Gathering information about users' preferences of lockers can improve the estimations of the PoU. For example, if a locker is less likely to be selected due to its distance from the station's kiosk, there is a larger probability that a usable bicycle will be parked there for a long period of time. On the other hand, if a preferred locker in which a bicycle is parked is not selected, it is more likely that the bicycle may be unusable.

Next, we introduce additional notation needed in order to incorporate user locker preferences in the model:

l	Locker id, $l \in C$
$\mathbb{L}(i)$	The locker in which bicycle i is parked
$\mathbb{L}(S)$	The set of lockers in which the set of bicycles S are parked
$a(l, L)$	The probability that locker l will be selected from within the set of lockers $L \subseteq C$

The values of locker selection probabilities $a(l, L)$ satisfy $\sum_{l \in L} a(l, L) = 1 \, \forall L$ and $a(l, L) = 0 \, \forall L \forall l \in C \setminus L$. The function $a(l, L)$ may be defined explicitly for each subset of lockers L or implicitly by some oracle that is capable of calculating or estimating it. Eqs. (3) and (4) can be re-written to accommodate users' locker preferences as follows, respectively:

$$P^e(j\ rented) = \left(1 - p_j^{e-1}\right) \cdot \sum_{S \in F(S^e \setminus \{j\})} a(\mathbb{L}(j), \mathbb{L}(S \cup \{j\}) \cdot \prod_{k \in S} (1 - p_k^{e-1}) \prod_{k \in S^e \setminus (S \cup \{j\})} p_k^{e-1}$$

$$P^e(i\ unusable,\ j\ rented) = p_i^{e-1} \cdot \left(1 - p_j^{e-1}\right) \cdot \sum_{S \in F(S^e \setminus \{i,j\})} a(\mathbb{L}(j), \mathbb{L}(S \cup \{j\}) \cdot \prod_{k \in S} (1 - p_k^{e-1}) \prod_{k \in S^e \setminus (S \cup \{i,j\})} p_k^{e-1}$$

Where $F(S)$ denotes the collection of all subsets of set S. Note that these equations are more complex as compared to (3) and (4), since not only the number of usable bicycles that are parked in the station is taken into account, but also the location of these bicycles.

Similarly, the approximated conditional probability (5) can be updated as follows:

$$\tilde{P}^e(j\ \text{rented}|j\ \text{usable}) = \frac{a(\mathbb{L}(j), \mathbb{L}(S^e))}{a(\mathbb{L}(j), \mathbb{L}(S^e)) + \sum_{k \in S^e \setminus \{j\}} a(\mathbb{L}(k), \mathbb{L}(S^e)) \cdot \left(1 - \tilde{p}_k^{e-1}\right)} \tag{11}$$

Note that if $a(\mathbb{L}(i), \mathbb{L}(S^e)) = \frac{1}{|S^e|} \, \forall i \in S^e$, i.e. all lockers have the same probability to be selected, Eq. (11) is reduced back to Eq. (5). In addition, given that all the bicycles in the station are usable, the probability in Eq. (11) equals $a(\mathbb{L}(j), L^e)$, namely, the probability that locker $\mathbb{L}(j)$ will be selected.

Due to the same mathematical arguments as in Section 3, we obtain the following iterative equation for updating the PoU of bicycle i after event e:

$$\tilde{p}_i^e = \tilde{P}^e(i\ unusable|\ j\ rented\ ) = \tilde{p}_i^{e-1} \cdot \frac{a(\mathbb{L}(j), \mathbb{L}(S^e)) + \sum_{k \in S^e \setminus \{j\}} a(\mathbb{L}(k), \mathbb{L}(S^e)) \cdot \left(1 - \tilde{p}_k^{e-1}\right)}{a(\mathbb{L}(j), \mathbb{L}(S^e \setminus \{i\})) + \sum_{k \in S^e \setminus \{i,j\}} a(\mathbb{L}(k), \mathbb{L}(S^e \setminus \{i\})) \cdot \left(1 - \tilde{p}_k^{e-1}\right)}$$

Note that if $a(\mathbb{L}(i), \mathbb{L}(S^e))$ equals 0, we obtain $\tilde{p}_i^e = \tilde{p}_i^{e-1}$. In other words, if the probability that a locker $\mathbb{L}(i)$ will be selected equals zero or is close to zero, the fact that bicycle i was not selected in event e does not provide information regarding the usability of bicycle i.

5.2. Station idle time

Until now we considered methods to update the PoU only at rent events. However, there may be situations in which for relatively long periods of time no rent event occurs in a station, even though bicycles are parked in it. This may be explained by one of the following: (1) no renters have arrived at the station (2) renters have arrived at the station but none of the bicycles were in a usable condition and no rent transaction has occurred (3) the station is malfunctioning. Recall that in the information system, there is no direct evidence for any of these occurrences. Here we present a method to update the PoU between rent events in order to call the attention of the operators to situations (2) or (3).

The arrival rates of renters during different time periods along the day can be estimated using trip history transaction data. If no rent event occurs for a long period of time in a non-empty station even though the estimated arrival rate of renters is high, the probability that the parked bicycles in the station are unusable (or cannot be rented due to station failure) increases.

Here we assume that the demand for bicycles is a time heterogeneous Poisson process. We denote by T the elapsed time since the last rent event in a station, and let $t_1, t_2, \ldots t_m$ ($T = \sum_{r=1}^{m} t_r$) be the lengths of consecutive time intervals. The expected number of arrivals of renters at each of these time intervals is denoted by $\mu_1, \mu_2, \ldots, \mu_m$, then the probability that no renter arrived until time T is:

$$\prod_{r=1}^{m} \exp(-\mu_r t_r) = \exp\left(-\sum_{r=1}^{m} \mu_r t_r\right)$$

Given that the elapsed time since the last rent event (e) is T, the PoU of bicycle i is recalculated as follows. If no renter arrived at the station, the PoU of bicycle i is $\tilde{p}_i^e$. However, if one or more renters arrived at the station but no rent event occurred, then bicycle i is unusable. By conditioning over these two complementary events and multiplying by their corresponding probabilities we can update the PoU of any bicycle i in the station to:

$$\tilde{p}_i^e \cdot \exp\left(-\sum_{r=1}^{m} \mu_r t_r\right) + 1 \cdot \left(1 - \exp\left(-\sum_{r=1}^{m} \mu_r t_r\right)\right)$$

Note that this updating expression depends on the time in which it is performed due to the dependency of $\mu_1, \mu_2, \ldots, \mu_m$ on this time. Noticeably, the PoU increases with the arrival rates in the given time intervals and the length of these time intervals. This update is effective until the next rent event at the station occurs. Once a rent event occurs the PoU is updated as discussed above in Section 3 or as in Section 5.1.

5.3. Enhancing the estimation of the prior probabilities

In this section, we discuss additional available information that can be used to estimate the prior probabilities. A generic estimator for the prior probability may be obtained by dividing the total number of bicycles repaired in a given period by the total number of the trips taken in the same period. However, the prior of each specific bicycle can be better estimated given data features such as: elapsed time since its last repair, accumulated riding time, mileage, usage areas, users' characteristics, etc. Specifically, it might be reasonable to assume that the prior probability of a bicycle that is returned from maintenance is close to zero. For a discussion on classes of life distributions based on notions of aging, see [1].

In addition to user's trips and maintenance activities, bicycles may also be removed from a station for the purpose of rebalancing the stations' bicycle inventory levels (repositioning activities). If the repositioning worker is instructed to check the condition of each loaded/unloaded bicycle, we can assume that when the bicycle is returned to a station at the end of the repositioning its prior probability to be unusable is close to zero. Alternatively, the calculation of the conditional probability may continue from the calculated value right before the repositioning.

Other aspects that can be taken into account when estimating the prior probability are the transactions' characteristics. For example, a short time (less than two minutes) round-trip (identical start and end stations) transaction suggests that a user unlocked a bicycle from a locker and almost immediately returned it to the same station. This may indicate that the bicycle is unusable. This kind of transaction is not rare; one may evaluate the percentage of times this kind of transaction was followed by a maintenance activity. This can be done by cross-checking transaction history and maintenance data.

Failure reports provided by users, i.e., by complaint calls or by a maintenance button, installed on the locker, can also be incorporated into the model. In particular, if user complaints are considered highly reliable, the reported bicycles can be flagged as unusable, i.e. $p_i^e = 1$. Given such information, the unusable bicycle can be removed from the set of available bicycles in the station, and the PoU of the other bicycles can then be updated accordingly.

5.4. Locker failure detection

Another failure type that may decrease the quality of service is locker failures. Specifically, the electro-mechanical locking system may sometimes fail to work properly. When this occurs, the users cannot rent or return the bicycles at such lockers. If the locker is occupied with a bicycle, this bicycle will eventually be flagged as unusable using our method. However, if the faulty locker is vacant it will be left empty until the locking mechanism is repaired. Such type of failure is not reported in the information system, and so the online state of the stations presented to the users may not be accurate. Currently, the operators cannot remotely detect such failures.

A complementary model equivalent to the one presented in Sections 2 and 3 may be formulated in order to assess the usability of vacant lockers. As a mirror scenario, the data to be used are the returning transactions. On each return of a bicycle to a station, we calculate the conditional probability that a locker is unusable given that bicycles were not returned to it. And again, as more return events occur in a station there is a greater probability that a locker that is left empty is unusable.

6. Discussion

In this paper, we presented a method to detect unusable resources (bicycles and lockers). Our approach introduces new real-time estimation of the number of unusable bicycles and lockers, which is currently not available to the operators. We validated our detection model using a simulation model that is based on data from a real-world system and demonstrated that it predicts well, in real-time, the number of unusable bicycles in a station. This is achieved without any knowledge about the arbitrary stochastic process according to which unusable bicycles arrive in a station. In addition, we presented several extensions of the model that may enhance the quality of the model's predictions based on more detailed data about the process that is available to the operators.

As discussed in Section 4, a further validation of the detection model can be accomplished using real-world data regarding unusable bicycles and lockers. We call practitioners to collect and make use of such data in their planning process. Retroactively, this data can be used to continuously fine-tune the detection model.

One limitation of our model stems from the assumption that bicycle failure is a binary property, i.e., a bicycle is either usable or unusable. In reality, some bicycles that require maintenance due to minor failures but may still be rented by the users. Bicycles in such a condition cannot be detected by the proposed model. A different detection model that considers the long-term transaction history of the bicycles in the system can be devised to detect such failures.

The negative implication of the presence of unusable resources is the reduction of the station capacity and the presentation of misleading information to the users. Unusable bicycles/lockers may have different effect on the quality of service in different stations, depending on the capacity of the station and the demand patterns. Evaluating this effect may assist in prioritizing the stations that should be visited by maintenance and repositioning workers. Once this is determined, the next planning stage is to determine the routes of the maintenance workers.

Acknowledgments

We thank Jacob Doctoroff and Chris Lewis from NYC Bike Share (CitiBike) for sharing with us valuable information and knowledge.

References

[1] Barlow RE, Proschan F. Statistical theory of reliability and life testing: probability models. Tallahassee: Florida State Univ.; 1975.
[2] Dell'Amico M, Hadjicostantinou E, Iori M, Novellani S. The bike sharing rebalancing problem: mathematical formulations and benchmark instances. Omega 2014;45:7–19.
[3] Forma IA, Raviv T, Tzur M. A 3-step math heuristic for the static repositioning problem in bike-sharing systems. Transportation Research Part B: Methodological 2015;71:230–247.
[4] Johnson NL, Kotz S, Kemp AW. Univariate discrete distributions. New York: John Wiley & Sons; 1992.
[5] Kaspi M, Raviv T, Tzur M. Bike sharing systems: user dissatisfaction in the presence of unusable bicycles (Working paper). Tel Aviv, Israel: Tel-Aviv University; 2015.
[6] Kaspi M, Raviv T, Tzur M. Collection of unusable bicycles in bike-sharing systems (Working paper). Tel Aviv, Israel: Tel-Aviv University; 2015.
[7] Lewis C. Bicycle redistribution operations manager at NYC bike share, personal communication, March 2015.
[8] NYC Bike Share August 2014 Monthly Report, ⟨http://www.citibikenyc.com/system-data/operating-reports⟩.