ARTICLE IN PRESS

Omega ∎ (∎∎∎∎) ∎∎∎–∎∎∎

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

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/

Please cite this article as: Kaspi M, et al. Detection of unusable bicycles in bike-sharing systems. Omega (2016), http://dx.doi.org/10.1016/j.omega.2015.12.003

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 $P^e(i\ usable|\ i\ rented)=1$ and $P^e(i\ rented|\ i\ unusable)=0$. 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:

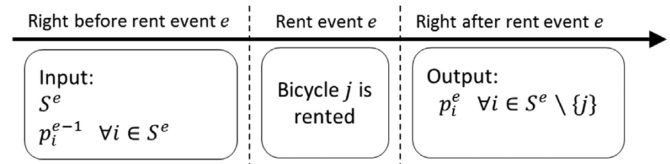

Fig. 1. Updating the PoU at rent event e.

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

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:

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

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: $P^e(j\ usable\ |\ j\ rented)=1$, 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:

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

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

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:

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

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:

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