Bike Sharing in Fakeville

Shane G. Henderson, inspired by ideas presented by Robert Hampshire on real bike sharing systems at a

March 31, 2010

(Integer-ordered variables, constrained.)

Problem Statement: In rough terms the problem is how to allocate a fixed number of bikes to stations in order to minimize the long-run costs of operating a bike-sharing program.

Bikes Inc has been contracted to run a bike-sharing operation in the city of Fakeville. The system operates as follows. A number s of stations are maintained around the city. At each station there are spaces in racks to store bikes. Riders swipe their membership card, thereby releasing a bike, ride the bike to a station near their destination, and return the bike to the racks there. If there are no bikes available when a rider arrives at a station then that rider is "lost," i.e., finds some other way to reach their destination. If there are no spaces available in the racks at the destination station, the rider waits until a space becomes available. If the rider is still waiting at the end of the day, then we will assume for simplicity that the bike is somehow left securely at that station. (This avoids the complication of trying to model user behavior when racks are full.)

Bikes Inc incurs a penalty p_e for each unit of time that each station is empty (no bikes), and a penalty p_f for each unit of time that a station is full (no space in the racks).

At the end of each day, Bikes Inc rebalances the system as follows. Let x_i (respectively q_i) be the target (actual) number of bikes at Station i at the end of each day, for each i. Thus, x_i is also the target number of bikes at Station i at the start of the following day. Bikes Inc redistributes the bikes so that there are, indeed, x_i bikes at Station i for each i. Ideally one might solve some form of trans-shipment or vehicle-routing optimization problem to determine the costs of this rebalancing, but for simplicity we instead assume the following rebalancing scheme.

Consider the stations in order from 1 up to s. For each station with a surplus ($q_i > x_i$), allocate the excess to the lowest-numbered stations with a shortage ($q_i < x_i$), until each station i has x_i bikes. (We ignore the complication of missing or nonfunctional bikes.) The cost of this redistribution is given by $R(q, x) = r \sum_{i,j=1}^{s} d_{ij} y_{ij}$, where d_{ij} is the Manhattan distance from Station i to Station j in kilometres (km), y_{ij} is the number of bikes that are moved from i to j, and r is a parameter.

The goal is to choose x to minimize the expected daily costs. More precisely, let $Q_i(t)$ denote the number of bikes in the racks at Station i at time t, and suppose the day runs from time $t = 0$ to time $t = T$. Further, let c_i be the number of bikes that can be stored in racks at Station i (the capacity of Station i), and let $I(\cdot)$ be the indicator function that is 1 if its argument is true and 0 otherwise. Then the goal is to minimize

$$E\left[p_e \sum_{i=1}^{s} \int_0^T I(Q_i(t) = 0)\, dt + p_f \sum_{i=1}^{s} \int_0^T I(Q_i(t) = c_i)\, dt + R(Q(T), x)\right]$$

over choices of the vector x that further satisfy $x \leq c$ (componentwise) and $\sum_i x_i = b$, where b is the total number of bicycles. Here $R(q, x)$ is the cost of rebalancing the bikes at the end of each day. (Credit for this optimization formulation is due to Robert Hampshire.)

Some further information on the problem is as follows.

1. In reality, the arrival process of users to stations is heavily time dependent, but we ignore that complexity in our formulation. We instead assume that users arrive to Station i according to a Poisson process with rate λ_i users per hour, and the arrival processes at different stations are mutually independent.

2. From Station i, users select their destination station according to the probabilities $(p_{ij} : 1 \leq j \leq s)$, where $\sum_j p_{ij} = 1$ for each $i = 1, 2, \ldots, s$.

3. The time from when a user borrows a bike at Station i to when they arrive at Station j to return the bike is distributed as a gamma random variable with mean $20d_{ij}$ minutes and variance $25d_{ij}$ minutes^2, and trip durations are independent of each other and of all other quantities. The only exception is when $i = j$, i.e., the bike is returned to the station from which it was taken, in which case assume a gamma distributed time with a mean of 45 minutes and a variance of 49 minutes.

Recommended Parameter Settings: Representative but not real data is as follows. The city is laid out as a square with sides of length 5km. Take the day length $T = 16$ hours. Assume $b = 3200$ bikes and $s = 225$ stations. Take $p_e = p_f = \$50$ per station per hour and the proportionality constant for rebalancing $r = \$5$ per bike per km. Further (generated) data is contained in an attached file.

Starting Solutions: Let

$$z_i = \min\{\lfloor b/s \rfloor, c_i\}$$

be an initial allocation of bikes to stations that respects the station capacities but is otherwise an equal allocation. If there are any remaining bikes then allocate them sequentially at random, in proportion to the number of free spaces in the racks at each station. For example, if there are 3 stations with capacities $c = (2, 4, 5)$, and $b = 10$ bikes, then we first allocate 2 bikes to Station 1, and 3 bikes each to Stations 2, 3. We assign the next bike (the 9th bike) to Station 2 with probability $1/3$ and to Station 3 with probability $2/3$. If the 9th bike goes to Station 2, then the remaining bike goes to Station 3. If the 9th bike goes to Station 3, then the remaining bike is assigned to Station 2 or Station 3 with equal probability. To generate multiple starting solutions, this process can be repeated independently.

Measurement of Time: One day's operation is one unit of time.

Optimal Solutions: Unknown.

Known Structure: None.