Cambridge Bluebike Distribution Optimization - Final Report

Team Members

Peijun Xu
Jiani He

Faculty Mentor

Dimitris Bertsimas
Dick Den Hertog

Student Mentor

Adam Cheol Woo Kim
Berk Ozturk

Date of Submission: 05/14/2021

Table of Contents

1 Problem Overview (0.5)

Bluebike is one of the leading bike-sharing firms in Boston, with over 3,000+ bikes and more than 300+ stations across Boston. However, it might be challenging for the bike service company to serve customers due to the imbalanced distribution of bikes among different stations and uncertainty in demand. It is common to have surplus bikes placed in one station and scarce bikes in another nearby station. Therefore, we want to propose a method to design a reliable Bluebike distribution system mainly in Cambridge. Also, because we are optimizing the number of bikes transferred, the nature of the problem involves integer decision variables. Although we touched on integer robust optimization in the class but didn't discuss extensively how to deal with robust optimization problems with integer decision variables, we would like to take this opportunity to explore efficient methods to tackle a multi-stage integer robust optimization problem and see if these techniques achieve similar effects as their applications on non-integer problems discussed in the class.

To save on computation time and achieve a realistic possible route, all 8 Cambridge stations, as the top 10 most popular stations are chosen as the region of interest, using real data from the Bluebike system in 2018-2019. We further aggregate the trip data into daily level and formulate the problem with daily-level decision variables. For example, although Bluebike staff transfer the bikes from station i to station j several times a day in real life, we make the assumption that they would only make the transfer once.

2 Data

2.1 Data Source

The data we use for modeling mainly come from the official Bluebike website, which provides the following data:

1. Trip History Data
    a. We use it to predict the number of rides starting from the station (demand) and number of rides stopping at the station (supply)
    b. The uncertainty in demand is estimated based on the error made by the prediction model on the historical data
2. Station Data
    a. Number of docks at each station
    b. We estimate the transfer cost as well as the correlated uncertainty, based on the distances between stations calculated using the coordinates.
3. Membership and Ridership

    a. We estimate the opportunity cost of unmet demand based on the how Bluebike charges the customers
4. Hiring information
    a. We estimate the cost of “emergent transfer” if the number of bikes at the station exceeds its capacity based on the salary offered to the staff

2.2 Cost Estimation

2.2.1 Transfer Cost

Bluebike hires bike technicians to move the bikes between stations. The daily salary for a bike transferer is $12; we assume he can move 30 bikes at 1 time from station i to station j. We also assume that he carries the bike from station i to station j only twice a day. Therefore the average cost to move one bike through bike technicians would be 12/(30*2)=$0.2 per bike.
There are 3 types of transfer and each incurs a different cost.

1. Internal Transfer: Transfer between Cambridge stations, cost will vary by the distance of the transportation.
2. External Transfer: Either transfer from a station outside Cambridge or transfer to a station outside Cambridge. We set the external moving cost as a fixed value 2 times the max internal transfer cost.
3. Emergent Transfer: Emergent transfer is not a part of the decision variables, it occurs only when the number of bikes at stations approaches 0 or the full capacity. We set it as a value higher than external transfer cost.

2.2.2 Mismanagement Cost

Mismanagement cost occurs in two situations.

1. If the decisions made are unable to meet all the demand, there would be opportunity cost as well as cost of emergent transfer
2. If the decisions made caused the number of bikes exceeds the station’s capacity, there would be cost of emergent transfer

2.3 Demand Estimation

While demand for bikes and stations depends on factors such as the time of day and weather, holiday effects, our focus is not on the demand prediction. Therefore, to generate our unsatisfied demands from station i to station j, we will build a time-series prediction model through the 'Auto Regressive Integrated Moving Average’ (ARIMA) with cross validation to select the hyperparameters and set the seasonality = 7 to capture the weekly seasonality. We use the daily net change(demand-supply) of each Cambridge station in 2019/01/01 - 2019/09/30 to predict the daily net change(demand-supply) of each Cambridge station in 2019/10/01 - 2019/12/31.

2.4 Uncertainty

We estimate the P matrix based on the ARIMA model's performance on 2018's data and the distance between each station.

3 Methods (1)

3.1 Robust Optimization

Parameters

We first defined the following parameters to incorporate into our model:

网图可以参考

- C: Capacity of the truck.
- $I_i = [\lambda_i - \omega_i, \lambda_i + \omega_i]$: The ideal threshold of bikes at station i after rebalancing, which is represented by a predetermined idea amount λ_i associated with a symmetric threshold of $2\omega_i$.
- γ_i: Capacity (number of docks) of bike station i.
- e_i: The existing number of bikes at station i before rebalancing.
- $d_{i,j}$: The Manhattan distance between station i and station j.
- n: Number of stations.
- w: Wage for driver, dollars per bike.
- h: Disinfecting cost, dollars per bike.
- g: Price of Gasoline, dollars per kilometer.
- s_i^+: A binary value that indicates whether there is a bike surplus at station i, i.e. $s_i^+ = 1$ when $e_i > \lambda_i + \omega_i$, and zero otherwise.
- s_i^-: A binary value that indicates whether there is a bike shortage at station i, i.e. $s_i^- = 1$ when $e_i < \lambda_i - \omega_i$, and zero otherwise.

Decision Variables

Next, we defined our decision variables. These variables define the choices the truck driver must make, specifically whether the driver will visit a station or not, and how many bikes they will pick up or drop off.

$x_{i,j}$: A binary variable that indicates the relocation truck passes through station i to station j if $x_{i,j} = 1$, and zero otherwise.

c_i: The current number of bikes on the truck before arriving at station i.

d_i: The number of bikes the truck drops off at station i.

p_i: The number of bikes the truck picks up at station i.

Objective Function

Our overarching objective here is to minimize the total operational costs, which is composed of estimated bike disinfecting costs per bike (h), gas costs per kilometre (g) and a fixed hourly wage for the truck driver (w). Here, B stands for the number of bikes that are relocated, while Drepresents the total distance that the truck has traveled.

The goal of the project is minimize the total opportunity to meet all predicted demands plus demand uncertainty through technicians' internal and external transfers by determining:

- the number of bikes placed in station i, b_i , at the beginning time when t=1
- the number of the bikes transferred internally from station i to station j within Cambridge at time t, X_{ijt}, where i = j =1,2,3,...,8
- the number of the bikes transferred externally from Cambridge station i to other Boston station at time t, Y_{it}

Demand uncertainty will be captured by P^TD, where D is constrained by budget uncertainty with bounded support [-1,1] so that we are guaranteed to be feasible for 95% of uncertain outcomes. There are several constraints that we will consider:

- The bikes available at station i at time t should be smaller than the bikes available at station i at time t-1 - (the unsatisfied demand i at t-1 + the unsatisfied demand uncertainty) + net transfer at station t during time t-1 + slack variables for potential loss
- There are two types of slack variables. One slack variable $slack\_more_{i,t}$ is to capture the potential loss due to lack of bikes at station i at time t; another slack variable $slack\_less_{i,t}$ is to capture the potential loss due to the lack of free docks at station i at time t. Either $slack\_more_{i,t}$ or $slack\_less_{i,t}$ has to be 0 at the given time.
- The number of available bicycles at station i at time t, B_{it} , is not smaller than 0 and no larger than the total number of docks at station i, D_i .

3.1.1 Robust Counterpart

3.2 Adaptive Robust Integer Optimization

3.2.1 Integer Uncertainty Constraints - Network flow

We round P to be integer matrix so that the problem can be formed as a network flow.

3.2.2 Gurobi Model Setup Constraints - Time Limit

3.2.3 Cutting Plane Method

4 Results (0.5)

5 Conclusions and Future Work

Add the influence of weather into prediction
ML to predict residuals
Subscriber or customer - customer analysis
Better prediction model, better data estimation.

6 Reference

7 Appendix

Formulations + variable definitions (2)
Visualization (2)