

Optimizing Hubway Bicycle Station Locations in Boston Area

Congyang Wang, Fangyu Lin, Meng Li

Department of Computer Science, Department of Data Science – Worcester Polytechnic Institute, MA, 01609

Email: {cwang8, flin, mli6}@wpi.edu

Abstract—As the number of people living in Boston grows, the traffic environment has become more and more crowded. In recent years, local government has already built up the bicycle system to ameliorate traffic conditions and create a green environment. The idea is to design and implement a bicycle system that is accessible and convenient to the public by planning station locations satisfying the public needs. Citizens can rent a bike from one station and return it at another bike station close to their destinations. However, it is important and challenging for the government to arrange bicycle stations appropriately so that the system can benefit almost everyone and balance traffic flows in Boston as much as possible. Our approach is to arrange bicycle station locations such that maximizes the coverage of the bicycle system and increase penetration of bicycle usage, based on the hubway bike data in Boston between 2011 and 2013. Furthermore, since data of capacity of hubway locations in Boston area is unavailable, we will estimate it with historical data, which contains information on stations and every trip. Our goal is to find the best locations to mitigate traffic problems and create a green environment in Boston Area.

Index Terms—Bike-sharing program; Bike-station location; Location Optimize model;

I. INTRODUCTION

Sustainable living and energy conservation have been populous across the globe in recent decades since now people are aware of severe damage on environment inflicted by industrialization. To

promote renewable energy, a significant number of developed countries launched a shared bicycle program in the hope that it will facilitate people to switch to green transportation. The Hubway is such program launched by Boston government in 2011 with 61 stations and 600 bicycles initially. By the end of 2013, the system has been expanded to 130 stations and 1200 bicycles. However, the current Hubway system is designed mainly for entertainment purpose. Obviously, the bicycle system has far more benefits than just for fun. It even has high potentials to solve urban problems of oversized and overpopulated cities, such as traffic congestion.

To explore more potentials, a new bicycle facility plan has to be designed based on trip demands, regardless of trip purposes. As a consequence, the new system will be more efficient and attractive compared to the current one as well as produce more beneficial outcomes. In the first place, reasonable bicycle facility planning will increase the usability of the system and correspondingly reduce usage of private automobiles and emission of carbon dioxide. Plus, the decline in consumption of fossil fuels will help to conserve the non-renewable resources. Secondly, traffic congestion will be mitigated if people switch to bicycles from automobiles. For low-income residents, the Hubway system will provide them an alternative and healthy commute solution. For college students, cycling is also a cost-effective and flexible option which better fits their school schedules compared to the subway. Additionally, an efficient bicycle system is likely to ease mass transit pressures for a metropolis.

To tackle the problem, we are going to use a dataset of the Boston Hubway 2011 Trips And

Stations to determine station locations of the future Hubway system which is supposed to be efficient and meet the public trip demands.

Basically, we will employ a location-allocation model to select stations such that costs between demand locations and candidate stations are minimized, or coverage of stations is maximized [1]. The dataset will serve a goal of estimating the distribution of potential demands for the location-allocation model.

According to the "Wisconsin Bicycle Planning Guidance Handbook (WBPG)" [2], it is necessary to take both supplies and demands of shared bicycles into consideration in the shared bicycle facility planning. So we decided to perform analysis with two models: the demand-based model and supply-based model.

The research paper we have reviewed currently is the one [3] that analyzes the bike rental system using the P-Median (minimize-impedance solution), which is to minimize "weighted costs between demand points and solution facilities" to maximize coverage. Moreover, in the paper, it uses the "GIS-based multi-criteria analysis tools." As we know, this is a popular tool for data visualization on a map creating a heat map rendering with other features. However, the tool is inadequate and flawed in certain aspects considering our problem. So we are going to apply a more advanced method. To estimate the capacity of each bike station, first, aggregate the number of bicycles rented from one station on a daily basis. Then select the maximum over daily total from each station as the estimated capacity. Regarding accessibility and regional population, some stations will have more bicycles that are rented than others. Based on the estimated needs, we can make a right decision to allocate bikes to stations to maximize overall usage of bicycles. Besides, we are reviewing other methodologies, such as multi-criteria decision analysis and exploratory spatial data analysis. We will make comparisons across them to figure out the best one for our problem and evaluate performance by comparing it with baseline algorithms.

II. OVERVIEW

III. METHODOLOGY

A. Problem Setting

Technically, we intend to solve two specific problems:

- Predict the number of available bicycles at each station, i.e., the number of bicycles allocated to each station in the very beginning.
- Predict the individual capacity of each station, namely, the number of docks at each station.

Intuitively, the sum of available bicycles at one station and bicycles that has traveled to it should meet the bicycle demands of the station. On the other hand, each station also has to provide sufficient empty docks to accommodate bicycles that have traveled to the station from other stations. Thus, the number of available bicycles and the capacity of each station should be predicted subject to these constraints.

The prediction consists of two steps. In the first step, bicycle demands and dock demands are predicted per station. In the second step, the entire Boston Area is split into 35 regions by zip codes. Numbers of bikes and docks that will be allocated to each station are estimated based on a demand ratio between the stations in every single region. For example, suppose the region with zip code 02101 has two stations A and B . Station A accounts for 20% bicycle demands of the total regional demands and station B accounts for 80%. Then 20% bicycles of the total bicycles in the region will be allocated to the station A and the remaining 80% bicycles will be allocated to the station B .

To make notations in our methods simple and clear, following terms are defined.

$Outbound_s$: the number of bicycles outbound from the station s .

$Inbound_s$: the number of inbound bicycles to the station s .

Avl_s : the number of available bicycles at the station s in the beginning.

$Capacity_s$: the capacity of station s , i.e., the total number of docks at the station s .

To simplify the problems and optimize allocation of bicycles and docks with the lowest costs and the least efforts, we made several assumptions.

- The total numbers of bicycles and docks of the bike-sharing system are fixed and equal to the totals of the current Hubway system throughout the study.
- Given trip data with start time and end time, regional population is assumed to be redundant for trip demand forecasts and will not be considered in the study.

B. Bike Demand Prediction for Stations

Since demands of bicycles are different from station to station, a strategy that distributes bicycles uniformly across all stations as adopted by most of bike-sharing systems is unreasonable. It may create a surplus of bicycles for stations with low demands and a shortage of bicycles for stations with high demands. To balance bicycle supplies and demands across stations, the number of available bicycles, in theory, should satisfy the equation below,

$$Avl_s + Inbound_s \geq Outbound_s$$

To cope with a special case where a station has no inbound bicycles, the bike allocation strategy has to guarantee sufficient bike allocation to meet the highest bike demands of the station in practice, i.e.,

$$Avl_s \geq \max(Outbound_s)$$

C. Capacity Prediction for Stations

The dock demand is predicted subject to two constraints:

- Every inbound bicycle is assigned an empty dock at a station.
- A station has enough docks to park as many bicycles as demanded when the demands reach the highest level.

$$Capacity_s - Avl_s = Inbound_s - Outbound_s$$

The left-hand side of the equation is the number of empty docks, and the right-hand side is the number of bicycles that have to be parked at the

station. The rationale behind the equation is that each station must have adequate empty docks so that every bicycle at the station can be parked with a dock. Meanwhile, when a station reaches its full capacity, in other words, no docks are empty, it should satisfy the highest bike demands, i.e.,

$$Capacity_s \geq \max(Outbound_s)$$

In a special case when a station has no outbound bicycles, it should be able to accommodate all inbound bicycles and bicycles that are allocated to the station initially, i.e.,

$$Capacity_s \geq \max(Inbound_s) + Avl_s$$

D. Estimation of Bike Allocation and Dock Allocation

Since capacity and quantity of available bicycles at each station of the current Hubway System are known, and the totals of docks and bikes are assumed unchanging, the regional total of docks and bicycles can be obtained.

With predicted bike demands and dock demands of individual stations, a demand ratio of bikes and that of docks at a given station s can be calculated respectively as,

$$Ratio_s^{Bike} = \frac{Predicted Avl_s}{\sum_{i=1}^n Predicted Avl_i}$$

$$Ratio_s^{Dock} = \frac{Predicted Capacity_s}{\sum_{i=1}^n Predicted Capacity_i}$$

where n is the number of stations in the region.

For station s , bike allocation and dock allocation should follow

$$Estimated Avl_s = Ratio_s^{Bike} * \sum_{i=1}^n Avl_i$$

$$Estimated Capacity_s = Ratio_s^{Dock} * \sum_{i=1}^n Capacity_i$$

where Avl_i and $Capacity_i$ are data of the current system.

E. Time Window

Prediction results may change when different time horizon is chosen, since daily $\max(Inbound_s)$, weekly $\max(Inbound_s)$ and monthly $\max(Inbound_s)$ probably are not identical. In our analysis, data will be aggregated on a daily basis, a weekly basis, a monthly basis and a yearly basis. Also, comparisons of accuracy of results produced with four horizons will be performed.

F. Data

The analysis is conducted with three datasets: hubway_stations.csv, hubway_trips.csv, and the most important and the largest data file, stationstatus.csv. Key attributes are summarized below,

Attribute	Data Type	Description Of Usage
seq_id (sid)	Character (Primary key)	trip index, counting trips
hubway_id	Character (Reference key)	Classification for station count sum
start_station	Integer	Demand-based counter
end_station	Integer	Supply-based counter
start_date	Timestamp	Classification for day, week, month, year
end_date	Timestamp	Classification for day, week, month, year
zip_code	Character	Classification for regions

IV. EVALUATION

V. RELATED WORK

VI. CONCLUSIONS

REFERENCES

- [1] J. C. Garca-Palomares, J. Gutierrez, and M. Latorre, "Optimizing the location of stations in bike-sharing programs: A GIS approach," *Applied Geography*, vol. 35, no. 1, pp. 235-246, 2012.
- [2] Tom Huber, "WISCONSIN BICYCLE PLANNING GUIDANCE" Unpublished, 2003.
- [3] N. Baird and D. Kim, "Bicycle Facility Demand Analysis using GIS," *Spaces and Flows: An International Journal of Urban and ExtraUrban Studies*, vol. 1, no. 2, pp. 1-14, 2011.
- [4] G. Rybarczyk and C. Wu, "Bicycle facility planning using GIS and multi-criteria decision analysis," *Applied Geography*, vol. 30, no. 2, pp. 282-293, Apr. 2010.
- [5] I. Frade and A. Ribeiro, "Bicycle Sharing Systems Demand," *Procedia - Social and Behavioral Sciences*, vol. 111, pp. 518-527, Feb. 2014.