New York City's Citi Bike System Simulation

Saul Toscano Palmerin

Cornell University

Contents

1 Introduction

We consider a realistic problem, using a queuing simulation based on New York City's Citi Bike system, in which system users may remove an available bike from a station at one location within the city, and ride it to a station with an available dock in some other location within the city. The optimization problem that we consider is the allocation of a constrained number of bikes (6000) to available docks within the city at the start of rush hour, so as to minimize, in simulation, the expected number of potential trips in which the rider could not find an available bike at their preferred origination station, or could not find an available dock at their preferred destination station. We call such trips "negatively affected trips."

2 Simulation

We simulated in Python the demand of bike trips of a New York City's Bike System on any day from January 1st to December 31st between 7:00am and 11:00am. We used 329 actual bike stations, locations, and numbers of docks from the Citi Bike system, and estimated demand and average time for trips for every day in a year using publicly available data of the year 2014 from Citi Bike's website Citi (2015).

We simulate the demand for trips between each pair of bike stations on a day using an independent Poisson process, and trip times between pairs of stations follows an exponential distribution. If a potential trip's origination station has no available bikes, then that trip does not occur, and we increment our count of negatively affected trips. If a trip does occur, and its preferred destination station does not have an available dock, then we also increment our count of negatively affected trips, and the bike is returned to the closest bike station with available docks.

We divided the bike stations in 4 groups using k-nearest neighbors, and let x be the number of bikes in each group at 7:00 AM. We suppose that bikes are allocated uniformly among stations within a single group.

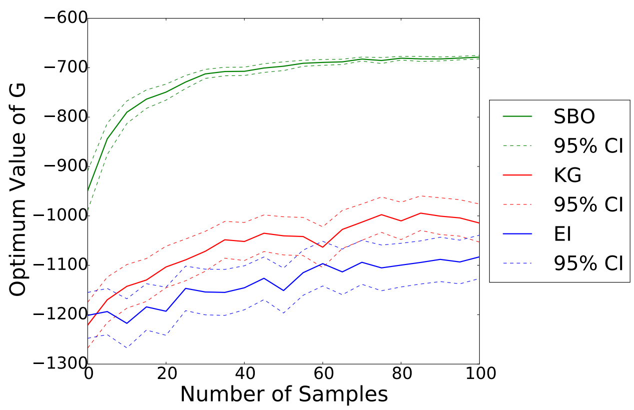
The simulation has five steps: (i) randomly choose a day i from January 1st to December 31st; (ii) simulate the total demand denoted by w as a Poisson random variable with mean λ_i, where λ_i is the total number of bike trips of day i; (iii) simulate the demands of the directed pairs of bike stations (j,k) as a multinomial random variable with parameters $\left(w, \left(p_{j,k} : 1 \leq j,k \leq 329\right)\right)$ where $p_{j,k}$ is the proportion of bike trips from from j to k on that day; (iv) simulate the arrival times to the bike stations as uniform random variables from 7:00am to 11:00am; (v) if a rider does not find an available bike, then he will leave the system, and we increment our count of negatively affected trips. If a rider does find an available bike, and his destination station does not have an available dock, then we also increment our count of negatively affected trips, and he will return the bike to the closest bike station with available docks.

3 Solution of the Citi Bike Problem

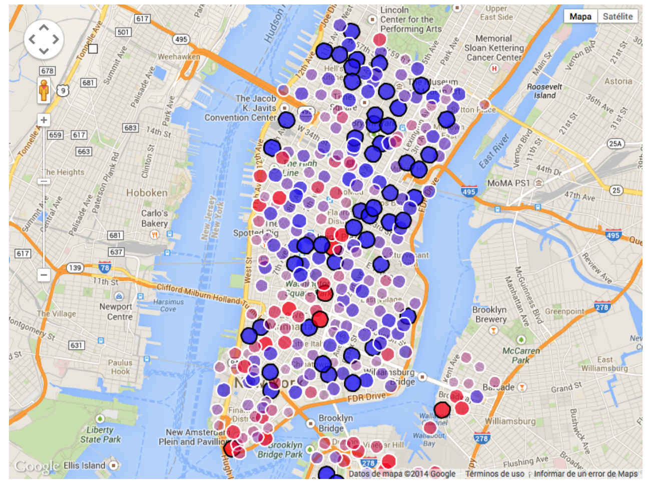
We solve the Citi Bike problem, arising in the design of the New York City's Citi Bike system, using three Bayesian optimization algorithms. We use Stratified Bayesian Optimization of (Toscano-Palmerin and Frazier, 2016), the Knowledge-Gradient policy of Frazier *et al.* (2009) and Expected Improvement criterion (Jones *et al.*, 1998).

We define our objective function by $G(x) = \mathbb{E}[f(x)]$ where $-f$ is the number of negatively affected trips between 7:00am to 11:00am, and x is the number of bikes in each group of bike stations at 7:00 AM.

Figure 1a compares the performance of Stratified Bayesian Optimization (SBO), Knowledge Gradient (KG) and Expected Improvement (EI), plotting the number of samples beyond the first stage on the x axis, and the average true quality of the solutions provided, $G(\text{argmax}_x \mathbb{E}_n[G(x)])$, averaging over 300 independent runs of the three algorithms. We see that SBO was able to quickly find an allocation of bikes to groups that attains a small expected number of negatively affected trips.



(a) Performance comparison between SBO and two Bayesian optimization benchmark, the KG and EI methods, on the Citi Bike Problem



(b) Location of bike stations (circles) in New York City, where size and color represent the ratio of available bikes to available docks.

Figure 1: Performance results for the Citi Bike problem (plot 1), and a screenshot from our simulation of the Citi Bike problem (plot b).

References

Citi (2015). Citi bike website. https://www.citibikenyc.com/, accessed November 2015.

Frazier, P., Powell, W., and Dayanik, S. (2009). The knowledge-gradient policy for correlated normal beliefs. *INFORMS journal on Computing*, **21**(4), 599–613.

Jones, D. R., Schonlau, M., and Welch, W. J. (1998). Efficient global optimization of expensive black-box functions. *Journal of Global optimization*, **13**(4), 455–492.

Toscano-Palmerin, S. and Frazier, P. (2016). Stratified Bayesian Optimization. *arxiv 1602.02338*.