Implementation of methods to predict demand, shortages, and evaluation of Resource allocation in the CitiBike System

Rohan Pitre(rp2816), Conrad de Peuter (cld2167)

1 Abstract / Introduction

Bike sharing systems are growing in popularity across the world. One main issue with bike sharing systems is station imbalance. This means that there are times when a rider goes to a station to pick up a bike and no stations are available, or that a rider goes to a station to return his bike and no docks are available. Data driven approaches are being used more and more to help the operations of bike sharing more efficient.

Currently, there is a lot of literature about predicting the demand of bike sharing systems and how to solve the optimization problems of rebalancing bikes at each station. However, these two methods haven't been applied simultaneously. Methods to solve the optimization problems assume a fixed demand rate which do not accurately represent how demand actually behaves. Furthermore, there are many heatmaps and dashboards depicting the current state of CitiBike, but none of these resources show predictions of future demand.

Our work will focus on integrating more realistic methods to predict demand in conjunction with quantitatively assessing strategies to rebalance the system. In particular, we will try to improve current methods to evaluate how balanced the system is by estimating out-of-stock events, as well as methods to assign routes to vehicles to move docks and bikes to have a more balanced system. Finally, we will explore approaches to encourage current riders to rebalance the system through promotions.

This problem is relevant to the course because we would like to build a system that can do this analysis in real-time. The current literature focuses on offline analysis and making these methods more efficient. Our goal is to integrate the current methods into one system that could potentially be used as a tool by bike-sharing systems to make decisions.

2 Data

2.1 Sources

2.2 Cleaning

3 Models

3.1 Station-based

3.2 Baseline

3.3 Neighborhood-based

4 Web Stack

5 Impact

6 Additional Resources

- Some computational time to run our optimizer algorithm to generate some query plans.
- Access to a machine where we can install and run experiments using our current database prototype.
- We will need to use some cloud computing infrastructure to complete this project. In addition we will need to have Spark Streaming working.

7 Related Literature

- *Background for the project:* Chen et al. introduce a Dynamic Cluster-based approach to predicting over-demand stations [**?**]. Li. et al. introduce a Gradient Boosting Regression tree model to predict the rate of demand going to and from an individual station. Zhang et al. introduce a regression model for individual trip destination and length predictions [?],. We plan on implementing these methods to find which are best at predicting demand in the system.

- *Work the project relies and builds on:* Some preliminary work has suggested a language for specifying query plans that we could borrow from. Also, Recent work nt ways to store and index data lend credence to the need for different cost models for access data in relations.

  Parikh et al. introduce a Markov Model to predict optimal restocking levels of a system [?]. Freund et al. consider the problem of optimal dock capacity [?].Raviv et al. also have an approach to determinig optimal inventory methods. We will use these approaches to evaluate how CitiBike restocks it's system.

  Schuijbroek et al. investigate optimal routing for system rebalancing in bike shares [?]. We plan on using their approach to evaluate the Citi Angels program.

8 Define Success

This project will be a success if we can accurately predict systems shortages ahead of time. After the models are fit, we will use them to predict real-time data. With better demand prediction we will be able to evaluate current resource allocation methods. If our model is able to correctly predict outages ahead of time, it will be a success. Finally, our web visualization will be useful for riders who expect to use the system, and the system operators to make strategic decisions by moving bikes and docks or offering incentives to riders .

References

[1] Zhang, J., Pan, X., Li, M., & Yu, P. S. (2016). Bicycle-Sharing System Analysis and Trip Prediction. 2016 17th IEEE International Conference on Mobile Data Management (MDM). doi:10.1109/mdm.2016.35

[2] Chen, L., Jakubowicz, J., Zhang, D., Wang, L., Yang, D., Ma, X., . . . Nguyen, T. (2016). Dynamic cluster-based over-demand prediction in bike sharing systems. Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '16. doi:10.1145/2971648.2971652

[3] Li, Y., Zheng, Y., Zhang, H., & Chen, L. (2015). Traffic Prediction in a Bike-Sharing System. Proceedings of the 23rd ACM International Conference on Advances in Geographical Information Systems.

[4] Freund, D., Henderson, S. G., & Shmoys, D. B. (2016). Minimizing Multimodular Functions and Allocating Capacity in Bike-Sharing Systems. arXiv preprint arXiv:1611.09304.

[5] Schuijbroek, J., Hampshire, R., & van Hoeve, W. J. (2013). Inventory rebalancing and vehicle routing in bike sharing systems. Chicago

[6] Parikh, P., & Ukkusuri, S. V. (2014, August). Estimation of Optimal Inventory Levels at Stations of a Bicycle Sharing System. In Transportation Research Board 94th Annual Meeting (No. 15-5170).

[7] Raviv, T., & Kolka, O. (2013). Optimal inventory management of a bike-sharing station. IIE Transactions, 45(10), 1077-1093.