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

Rohan Pitre, 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 One Sentence Summary

We will use methods to predict demand of CitiBike in New York City and use the results to improve rebalancing by using more realistic predictions. Our hypothesis is that incorporating predictions and operations in one system will improve efficiency.

3 Audience and Needs

This project will impact the academic community as well as decision makers of bike sharing systems. The community will benefit because our work will validate currently proposed methods to predict demand and allocate resources optimally.

Decision makers at CitiBike will benefit from having a fully integrated system that includes predictions of demand as well as suggestions on how to allocate resources. Additionally, we hope to create a dashboard visualizing our results in real-time which provides an interface for decision-makers to use. Finally, a major focus of this project is to design the architecture to make this analysis scalable.

4 Approach

1. We will implement methods to predict demand of riders that relies on historical data. These methods include basic time-series models (ARIMA) as well as more advanced methods that include clustering and gradient boosting trees.

2. We will evaluate these prediction methods to see which one provides the best predictions in an efficient manner.

3. With our predictive model in place, we will modify current methods to measure imbalance and optimize resource allocation to use our predicted demand rather than a fixed demand.

4. We will also evaluate the CitiBike bike angels program. This gives incentives to people who take trips to balance the system.

5. We will then build a dashboard to view both historical demand and live demand using CitiBike's live data stream.

5 (Best Case) Impact

In the best case scenario we are able to identify methods which accurately predict demand in the system. In addition we will be able to provide insight which may improve CitiBike's resource allocation. Finally, we build and publish online an all-inclusive dashboard which clearly shows all predictions and the effectiveness of rebalancing strategies both historically and in real-time.

6 Milestones

1. Train models using algorithms from literataure and CitiBike's open data to predict demand in the system, using historical data as validation.

2. Incorporating predicted demand levels into resource optimization models from literature we will evaluate CitiBike's current resource allocation methods

3. Set up Spark Streaming infrastructure using CitiBike's live feed, and use trained models in real-time predict areas where demand may not be met

4. Build dashboard showing historical and real-time system allocation

5. Evaluate CitiBike's angels program using methods from literature

7 Obstacles

7.1 Major obstacles

- A major obstacle would be if the models we use do not accurately predict system imbalances. If the prediction methods we implement from literature do not accurately predict demand, then we will not be able to go forward with the optimization evaluation. In this scenario we will still be able to build a live-web dashboard showing resource allocation across the city.

7.2 Minor obstacles

- We do not have any control over the restocking strategies CitiBike uses, thus we can only evaluate how they are doing compared to hypothetical system layouts we simulate. We are unable to see the true effectiveness of various restocking strategies from literature.

- Running a large scale optimization problem may require proprietary software.

8 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.

9 Literature Review

- *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.

10 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 who control the bike stocking.

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.