New York City Taxi Trip Supply and Demand Forecasting using TLC Trip Record Dataset Project Proposal

Kimberly Williamson, Gopal Menon
Data Mining
Spring 2017, University of Utah

I. Background

The New York City (NYC) Taxi and Limousine Commission (TLC) has released[1] yellow and green taxi trip data capturing pick-up and drop-off dates/times, pick-up and drop-off locations, trip distances, itemized fares, rate types, payment types, and driver-reported passenger counts. The latest data is the one that is available for the months January through June of 2016.

II. Project Scope

The project team will use the TLC 2016 dataset to forecast taxi trip supply, demand and expected fare for any given date, time and location using the regression methods that will be covered in the Data Mining 2017 Spring semester at the School of Computing of the University of Utah. The accuracy of the forecast will be evaluated using a metric based on the average forecast error on a test dataset. This metric will be developed by the project team. In addition to the techniques covered in class, the project team will also select two techniques that are either new or will be built on techniques covered in class with the aim of achieving a higher forecasting accuracy.

III. Project Team

Kimberly Williamson and Gopal Menon will work on the project.

IV. Data

The dataset will be downloaded from NYC TLC website and will be denormalized for the purpose of the project. In addition to this, local weather and event information will also be incorporated into the forecast.

V. Why this is interesting

The reason this problem is interesting is because it will provide a way for a New York City taxi customer to be negotiate a fair price for the taxi ride and also enable the taxi cab company to predict the number of rides that will be needed and in doing so reduce wastage of resources and also provide timely rides to customers. Forecasting of demand surges and drops will enable the taxi company to efficiently distribute resources.

VI. What is new

We plan to evaluate various linear and non-linear regression models that will result in the best prediction accuracy. At this point of time, we are not sure what new algorithms or techniques we will use in order for this project to have something new.

References

[1] "NYC Taxi & Limousine Commission - Trip Record Data." *NYC Taxi & Limousine Commission - Trip Record Data.* N.p., n.d. Web. 28 Jan. 2017.