A Cluster Analysis of Uber Request Data via Transit app in New York City

Jing Guo, Antora Mohsena Haque, Cassidy Crossland, Candace Brakewood

guo31@vols.utk.edu, ahaque3@vols.utk.edu, ccross3@vols.utk.edu, tblernew@utk.edu

Presenter: Jing Guo

1. Abstract
As ridehailing services like Uber become increasingly common in urban transportation systems, it is necessary to understand their usage patterns. Since private ridehailing companies do not publicly disclose their ridership data, usage patterns can be analyzed using other data sources, such as Transit app’s Uber Request data. The objectives of this research are three-fold: (1) explore the temporal characteristics of Uber requests through data visualization, (2) identify groups of users through cluster analysis, and (3) compare Trip-App Uber Request data with overall Uber usage data and transit data. The exploratory analysis results suggest that requests occurred most frequently during AM and PM peak periods. K-means clustering identified eight groups of Uber users: long duration and frequent users, long duration and infrequent users, party goers, long duration and infrequent users, holiday users, and weekend users. The main trip purposes determined by the clustering analysis were going to social events and to and from the workplace or home. Comparing Trip-App data to the overall ridehailing usage data and transit data suggest that the time distribution pattern of Uber app Uber requests is a combination of transit and Uber usage while the time usage features are more similar to those of Uber users. These results will help transportation departments to better coordinate ridehailing services and public transportation to meet users’ travel needs.

2. Objectives
1. Explore the temporal Uber requests pattern based on the time of day and day of week;
2. Infer trip purpose and identify distinct groups of users by K-means unsupervised machine learning algorithm; and
3. Compare temporal characteristics of Uber requests from Transit app with Uber pickup data released publicly through the TLC and General Transit Feed Specification (GTFS) service information for the subway system in New York City.

3. Methodology

4. Data
- **Dataset:** Uber requests obtained from Transit app
- **Study period:** 10.30.2016 to 10.29.2017
- **Study area:** New York City

5. Data preparation for clustering
- **Goal for data preparation:** Create new variables to explain the temporal usage patterns of Uber requests
- **New created variables:**
  1. Count of Days
  2. Duration
  3. AM Peak Usage Rate
  4. PM Peak Usage Rate
  5. Holiday Usage Rate
  6. Weekend Usage Rate

Transit app users’ Uber requests were mostly concentrated in the AM and PM peak periods during weekdays and late night periods on Fridays and Saturdays.

6. Exploratory analysis

7. Clustering analysis

8. Comparison with TLC Uber pickup data and transit

The time distribution characteristics of Uber requests via Transit app are likely a combination of transit usage and Uber usage, but the temporal features are more like those of Uber users.

ACKNOWLEDGMENTS:
The authors would like to thank TranApp for providing the data. We are particularly grateful to Jake Sion and David Block-Schachter for reviewing the paper.