Commuter classification and behavior clustering: Beijing use case

Selene Baez Santamaria
s.baezsantamaria@student.vu.nl

Abstract

Public transportation, centered on subway and bus networks, is an data-rich domain that can benefit from data mining and machine learning techniques. The classification of commuters versus non-commuter/occasional travelers can help government, transport management and operators to better target their policies in order to improve the transportation network in large cities. Furthermore, characterizing commuters by behavior clustering can bring deeper insight into their needs and routines as a whole. This project proposes the usage of ensemble models for classification and clustering of public transport users. For this purpose, transit card data will be used, available from the city of Beijing, China.

1 Introduction

1.1 Transportation domain

1.2 Beijing

1.3 Societal context

Public transport users

1.4 Scientific context

Usage of machine learning of data mining has been limited

Interdisciplinary study

2 Related work

Preprocess data

Machine learning for commuters identification

Ensemble methods

Classifiers in the transportation domain

3 Objective

3.1 Research questions

4 Methodology

5 Results

5.1 Commuters identification

Accuracy

Confusion matrix

5.2 Variable evaluation

5.2.1 Qualitative

Exploration: Experts opinion

5.2.2 Quantitative

Analysis: Correlation

5.3 Commuters clustering

5.3.1 Expert judgment

6 Conclusion

7 Future work

References

[1] Tu, Q. Weng, J. C. & Yuan, R. L. Impact Analysis of Public Transport Fare Adjustment on Travel Mode Choice for Travelers in Beijing. *CICTP 2016.*, pp. 850–863.