Handling Bias in Large Urban Datasets
Using Smart Fare Card Data for Transit Planning

Mark Hallenbeck, Mayuree Binjolkar, Daniel Dylewsky, Andrew Ju, Wenonah Zhang, Michael Wolf

CASCADIA URBAN ANALYTICS COOPERATIVE

Origin-Destination Data from Orca is Biased Against Short Duration Trips

ORCA DESTINATION
ORCA ORIGIN
ORCA TRANSFER

Example of ORCA Financial Transfers

ORCA can combine Two Trips into one
Trip A + Trip B = long trip with transfer

Research Task:
• To determine if a transfer was real or financial
• If transfer is financial then create new origin – destination pairs

Chose Semi-Supervised Based Machine Learning Approach for Classification of Types of Transfers

Labels learned from Label Spreading(KNN)

Outcome of Transfer Classification
Real Transfer
Financial Transfer

Transfer Classification Results by Stop

Many Real Transfers
Many False Transfers

Marker Size=Transfer Count
Number and Type of Transfer Activity

Statistical Bias Analysis

• Not all transit users are ORCA users
• Need to account for cash-paying customers for business planning
• Compare limited passenger count data to ORCA data

ORCA can combine Two Trips into one
Trip A + Trip B = long trip with transfer

Future Work

• Rebuilding origin-destination data files using transfer analysis results
• Deploy Hidden Markov Models for statistical bias analysis

Sponsoring Agencies