Transit Demand Analysis and User Classification Using Automatic Fare Collection (AFC) Data

Alireza Khani
University of Minnesota - Twin Cities

Let us know how access to this document benefits you.

Follow this and additional works at: https://pdxscholar.library.pdx.edu/trec_seminar

Part of the Transportation Commons, Transportation Engineering Commons, and the Urban Studies Commons

Recommended Citation
https://pdxscholar.library.pdx.edu/trec_seminar/144

This Book is brought to you for free and open access. It has been accepted for inclusion in TREC Friday Seminar Series by an authorized administrator of PDXScholar. For more information, please contact pdxscholar@pdx.edu.
Transit Demand Analysis and User Classification using Automatic Fare Collection (AFC) Data

Alireza Khani
Assistant Professor, Department of Civil, Environmental, and Geo-Engineering, University of Minnesota Twin Cities

Transportation Seminar, Portland State University, March 9, 2018
Transit Demand Analysis

• Crucial for service planning
  – Transit assignment: needs OD matrix, route choice behavior
  – Service frequency and timetabling: need spatial and temporal demand pattern, user perceptions, etc.

• Traditionally using on-board survey data
  – Small sample size
  – Every 5-10 years
  – Expensive to collect
  – Subject to errors
Transit Demand Analysis

• Automated transit data:
  – Very large samples
  – High resolution and detailed
  – More reliable measurements
  – Available every day

• Need new methods and tools
• Usually no user information available
Outline

• Transit Demand Analysis
• Automatically Collected Transit Data
• Demand Analysis Using AFC Data:
  – Descriptive analyses of demand
  – Origin-destination estimation using a trip chaining algorithm
  – User classification using trip chaining results
• Conclusions and Future Work
Automatically Collected Transit Data

- Automatic Vehicle Location (AVL) Data
  - GPS points of buses every few seconds

- Good for:
  - On-time performance analysis
  - Speed and delay analysis
  - Transfer reliability analysis
Automatically Collected Transit Data

• Automatic Passenger Count (APC) Data
  – Number of ONs and OFFs at each stop for each vehicle trip

• Good for:
  – Ridership analysis
  – Demand estimation
  – Model validation
Automatically Collected Transit Data

• Automatic Fare Collection (AFC) Data
  – Smart card TAG information (location, time, route, dir, etc.) for each passenger trip

• Good for:
  – Ridership analysis
  – Demand estimation
  – User behavior modeling
Demand Analysis using AFC Data
Case study on University of Minnesota student passes
U-Pass

- University of Minnesota students pass
- $100 per semester
- Unlimited ride in Metro Transit regional network
- Tag frequency declined since 2009

www.metrotransit.org/upass
U-Pass

• Objective:
  – Analyze changes in travel pattern of university students over time using U-Pass data
  – Cluster students according to their origin-destination and travel behavior

• Results and findings to be used for better marketing of U-Pass towards more transit usage by students
U-Pass Data

• Every time a user with U-Pass rides transit, the system records
  – Card ID
  – Tag time
  – Tag location
  – Route number
  – Transfer (2.5 hr free transfer)

• There is no information on
  – Origin-destination
  – Path
U-Pass Data – Descriptive Analysis

Tag frequency (ridership) per school year

~23% decrease

Tag frequency (ridership) per school year
U-Pass Data – Descriptive Analysis

Number of unique cards used per school year

~8% increase

Number of unique cards used per school year
U-Pass Data – Descriptive Analysis

Average tag per card per school year

~28% decrease

Average tag per card per school year
U-Pass Data – Descriptive Analysis

Ridership by day

Ride per card by day

Ridership by month

Cards used by month
U-Pass Data – Time Series Analysis

- Monthly ridership over six years

Decomposed

- Observed
- Trend
- Seasonal variations
- Random variations
U-Pass Data – Time Series Analysis

- Monthly unique cards used over six years
U-Pass Data – Some Findings

• U-Pass ridership does have a decreasing trend

• Number of cards used per year is picking up since 2014 (when Metro Green Line opened)

• Seasonal variations show that students buy the pass, but use it less towards the end of school year
Demand Analysis using AFC Data

Origin-destination estimation using a trip chaining algorithm
Trip Chaining – Concept

- Given tag locations and times, infer a chain of trips, paths, origin and destination of the user

- Assumptions:
  - Users start their first trip of the day from home and end their last trip of the day at home
  - During the day, they only use transit (no other mode)
  - Users start a trip near the end of the previous trip (do not walk for a long distance)
Trip Chaining – Method

• Overall algorithm:
  – Find the nearest stop to the tag location and mark it as boarding
  – Find the vehicle trip nearest in time to the tag time and mark it as the boarding time
  – Find the nearest stop to the next tag location and mark it as alighting, find the alighting time on the same trip
  – For the last trip of the day, use first tag as the next tag
Trip Chaining – Possible Issues

- Incorrect boarding stop inference due to GPS error
- Incorrect trip ID inference due to service delay
- Incorrect alighting stop inference due to incorrect trip ID (when routes have variations)
Trip Chaining – Proposed Algorithm

• Instead of inferring trip attributes sequentially, infer the most likely trajectory \((b, t, a)\) of the passenger

\[
P(b, t, a) = P_1(b). P_2(t|b). P_3(a|b, t)
\]

- \(P_1\): probability of boarding stop \(b\)
  - Determined by GPS error distribution
- \(P_2\): probability of trip \(t\) given boarding stop \(b\)
  - Determined by bus arrival delay distribution
- \(P_3\): probability of alighting stop \(a\) given boarding stop \(b\) and trip \(t\)
  - Determined by a route choice model, with the utility function including in-vehicle and walking time
Trip Chaining - Results

Initial data cleaning

<table>
<thead>
<tr>
<th>Description</th>
<th>Number of Tags</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total tags (Mar 7-10, 2016)</td>
<td>85,456</td>
<td></td>
</tr>
<tr>
<td>Tags with geographical coordinates issue</td>
<td>8,300</td>
<td>9.7%</td>
</tr>
<tr>
<td>Single tags</td>
<td>10,782</td>
<td>12.6%</td>
</tr>
<tr>
<td>Remaining tags</td>
<td>66,374</td>
<td>77.7%</td>
</tr>
</tbody>
</table>

Inference summary

<table>
<thead>
<tr>
<th>Tag Type</th>
<th>Number of Tags</th>
<th>Inferred (Baseline Algorithm)</th>
<th>Inferred (Proposed Algorithm)</th>
<th>Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regular</td>
<td>60,812</td>
<td>46,507</td>
<td>51,919</td>
<td>7%</td>
</tr>
<tr>
<td>Pay Exit</td>
<td>5,562</td>
<td>0</td>
<td>4,504</td>
<td>8%</td>
</tr>
<tr>
<td>Total</td>
<td>66,374</td>
<td>46,507</td>
<td>56,423</td>
<td>15%</td>
</tr>
</tbody>
</table>
Trip Chaining - Results

Morning origins

Morning destinations
Trip Chaining - Results

Metro Green Line Morning Trips

Eastbound

Westbound
Demand Analysis using AFC Data

User classification using trip chaining results
Spatial User Classification

• Representing changes in students’ origins (homes)
• Using origin destinations from trip chaining
• DBSCAN algorithm:
  – Does not fix the number of clusters
  – Needs the cluster radius
  – Needs minimum cluster members
Special User Clustering - Results

Feb 2009
Special User Clustering - Results

Feb 2012
Special User Clustering - Results

Feb 2016
Behavioral User Classification

• Using multiple days trip chaining results
• Representing user regularity in riding transit:
  – Number of days used transit
  – Average number of trips per day
  – Frequency of similar boarding stops
  – Frequency of similar routes
  – Frequency of similar departure time
• K-means algorithm to determine:
  – High-regular users
  – Mid-regular users
  – Low-regular users
Behavioral User Classification - Results

Share of cards in each cluster

Days traveled (out of 16)

Number of trips per day

Number of similar departure times

Legend:
- High
- Medium
- Low
Findings from Trip Chaining and User Classification

- Student riders became spatially less clustered by time, (more students live on or near campus and don’t use transit)
- Student riders became more regular in general:
  - High regular riders have kept using transit
  - Low regular riders dropped out
- Significant changes in travel patterns were observed in 2014, when Metro Green Line opened
How Can These Be Used?

• Metro Transit’s marketing strategies

• Fare structure and pricing of U-Pass

• Planning or adjusting service towards times and locations where there is more demand
Future Work

• Trip chaining algorithm could be improved by using AVL data instead of GTFS
• Extension to systemwide AFC data
• Regional OD matrix estimation
• Other clustering methods and attributes
• Route/stop choice modeling
Acknowledgments
Acknowledgments

Pramesh Kumar
PhD Student

Members of Transit Lab
http://umntransit.weebly.com/
Questions?

Alireza Khani
akhani@umn.edu