Ridership Patterns in an Urban Bike Share System

Hans Engler

June 12, 2015
... in New York City

Hans Engler

Bikeshare

June 12, 2015
... in Paris
Work with students since 2012

2012: Nathan Davis, Ryan McMillin, Michael Slattery (MS)

2013-14: Eric Buras \(^1\), Marcus Landers (undergrad)

Math-510 in 2012 and 2013

\(^1\)Honors thesis
Capital Bikeshare

- Started in 2010
- As of 06/15 there are 350 stations and 10,000 rides/day
- Detailed ride records are available
  - 0h 5m 41s, 6/30/2013 23:51, Florida Ave & R St NW, 6/30/2013 23:56, 5th & K St NW, W01380, Registered
- Current system status is also always available
Weekly Use Q2 2013

Hourly Ridership, April - June 2013

Riders

Monday  Tuesday  Wednesday  Thursday  Friday  Saturday  Sunday

Weekday

Subscribers
Casual Riders
An Hourly Station Status

12th & U St NW

Hans Engler
Bikeshare
June 12, 2015
Typical Trips

- Home ⟷ work, home ⟷ subway, subway ⟷ work, work ⟷ restaurant / club, restaurant / club ⟷ home
- Direct or multi-mode ("last mile")
- These occur at different times and between different stations
- Extract these temporal/spatial patterns
Trips between any station pair follow one of several temporal patterns.
Find these patterns and associate with station pairs.
300+ stations, $\approx 2 \cdot 10^5$ station pairs.
Assign to one of $O(1)$ clusters.
Use $O(10^6)$ rides in a given quarter.
Related Work

- A. Randriamanamihaga, E. Côme, L. Oukhellou, G. Govaert 2013
  *Work on Paris Velib‘ system that inspired this approach*

- E. O’Mahoney, D. Shmoys 2015
  *Optimization of rebalancing tasks in New York Citibike system*

- S. Thomas, Ph.D. Rice 2010
  *Clustering for time series of counts*
Casual riders and subscribers behave differently

- Weekdays and weekends (incl. Memorial Day, July 4, ...) are different

- Let’s use hour information of start of ride.

- Ride count vectors live in a 24-dim space

- 66 % of all station pairs never had a ride

- One station pair had \( \approx 400 \) rides/month
Hard clustering will just put the busiest station pairs into one cluster

Use soft model-based clustering

Poisson based model introduced by Govaert etc. for Paris Velib‘ system
Notation

- Station pairs $(i, j)$, time $t \in \{0, \ldots, T - 1\}$, cluster $\ell$
- $X_{ijt} =$ count of rides from station $i$ to $j$ starting at a time $\in [t, t + 1]$ during $D$ days of observation, $t = 0, 2, \ldots, T - 1$
- $Z_{ij\ell} = 1$ iff station pair $(i, j)$ is in cluster $\ell$, $Z_{ij\ell} = 0$ otherwise
Model

- \( Z_{ij} \sim \text{multinomial}(1, \pi_1, \ldots, \pi_L) \)
- \( X_{ijt} \mid Z_{ij\ell} = 1 \sim \text{Poisson}(D \cdot \alpha_{ij} \cdot \lambda_{\ell t}) \)
- \( X_{ij0} \perp \perp X_{ij1} \perp \perp \ldots \perp \perp X_{ij,T-1} \mid Z_{ij\ell} = 1 \)
- Normalization:  \( \sum_t \lambda_{\ell t} = T \)
- The \( \alpha_{ij} \) are mean ride counts from \( i \) to \( j \)
- The \( \lambda_{\ell t} \) are relative hourly intensities
- The \( Z_{ij\ell} \) are unobserved.
Approach and Implementation

- Compute parameter estimates $\hat{\alpha}_{ij}$, $\hat{\lambda}_{\ell t}$ and a posteriori probabilities $c_{ij\ell}$ of $(i, j)$ being in cluster $\ell$
- Use EM-algorithm
- The $\hat{\alpha}_{ij}$ can be found off-line
- Update equations can all be done with array operations in $\mathbb{R}$
Computational Performance

- One iteration takes $\approx 1$ sec on my cheap laptop
- Convergence after $O(100)$ iterations
- Clusters have distinct time patterns that are qualitatively reproducible
- The \textit{a posteriori} probabilities $c_{ij\ell}$ are $> .95$ for up to 80\% of all rides
Intensities $\sim$ # clusters

![Graphs showing intensity changes over time for different clusters.](image-url)
Intensities $\sim$ day $\times$ user

Registered riders on weekdays

Casual riders on weekdays

Registered riders on weekends

Casual riders on weekends
Morning, mid day, afternoon trips.
This is downtown, no nearby subway station.
Morning, mid day, afternoon trips.
On border of downtown and residential areas, subway and commuter rail.
Tunable parameters and variability

- Can select time window. *Patterns change with time!*
- Can select number of clusters. *3 – 6 clusters suffice, depending on location.*
- Variability comes from random initializations.
- Real and apparent change can come from system growth, seasonality, development of community preferences, new housing, new bars, new bus lines, *price hikes, new software, a string of bad accidents,* . . .
Conclusions

- Toolset to analyze ridership flow and its development
- Can be used to explore other bikeshare systems
- Can be used for system load predictions and simulations
- Holy grail: Describe multi-mode trips.