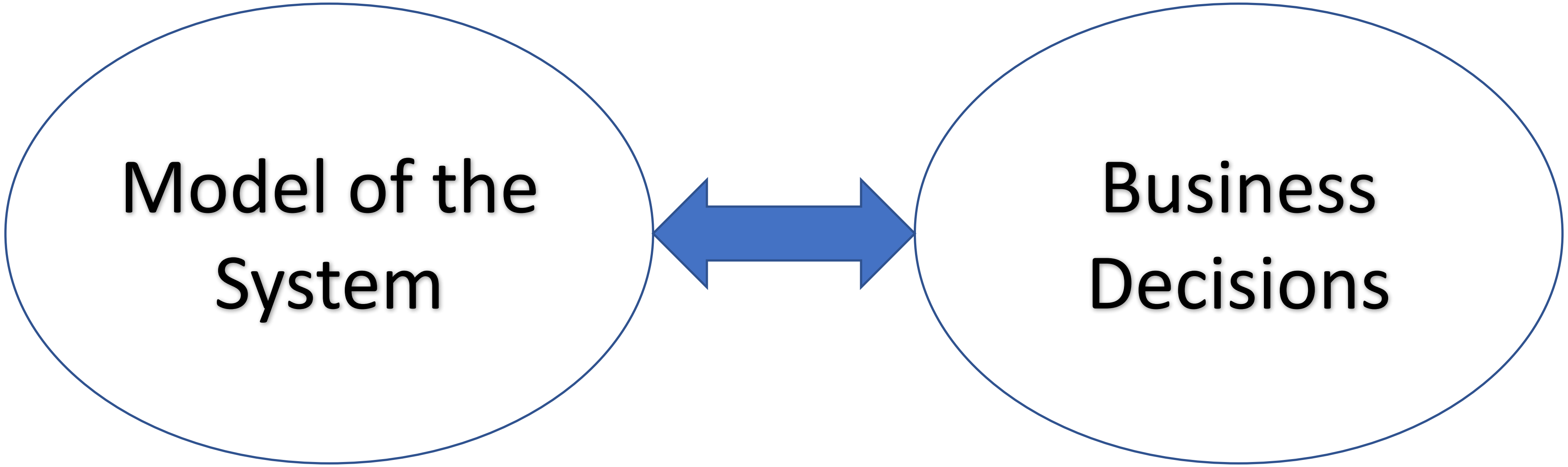
Models for Bike-Share Systems

Alice Paul

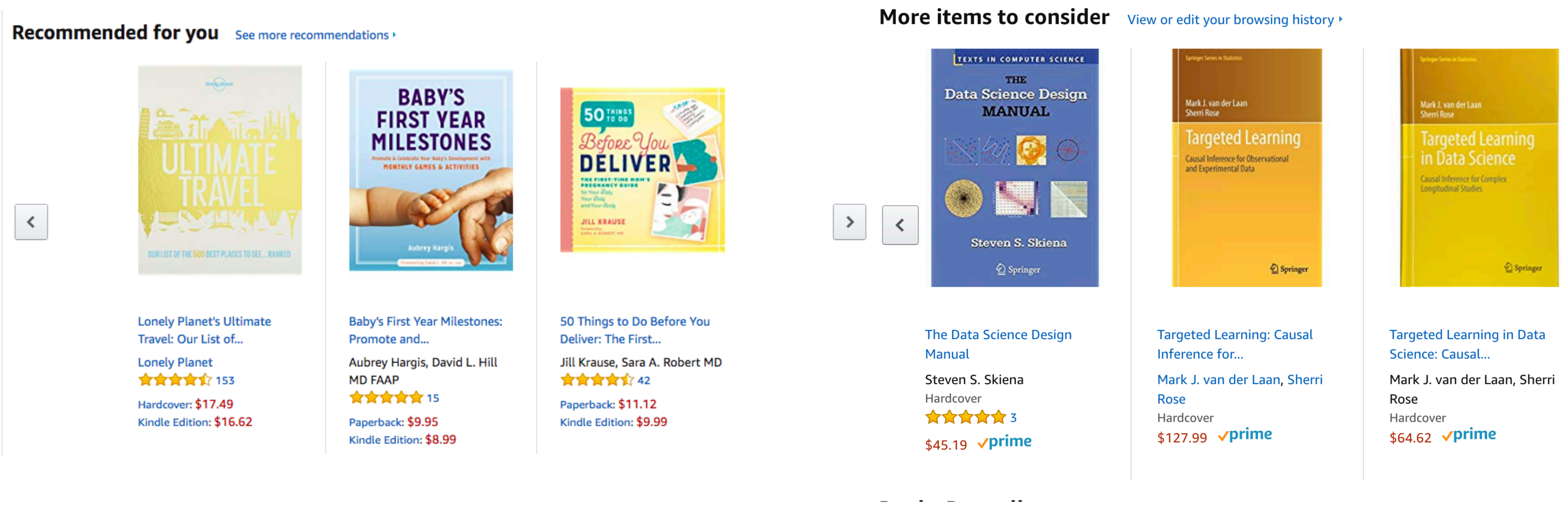
Care about Business Decisions



Disconnect between model and use

Alibaba Example

- Online retailer
- Want to predict purchases and choose which products to display



Alibaba Example

- Simple Multinomial Logit (MNL) model vs neural network to predict purchases among products.

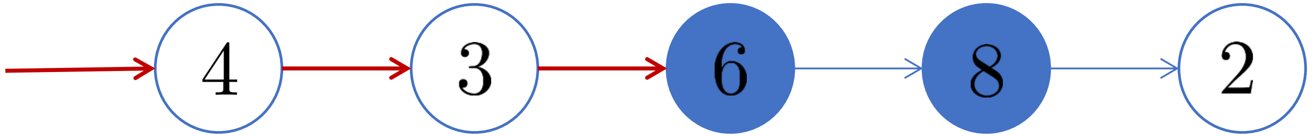
$$\Pr(buy\ product\ i) = \exp(\beta \cdot x_i)\ /\sum \exp(\beta \cdot x_j)$$

- Given the MNL model, we can easily optimize which products to display to maximize revenue because model is interpretable.
- Overall, neural networks do better at prediction but the MNL model produces more profitable assortments of products in A/B testing.

Feldman et al., Taking Assortment Optimization from Theory to Practice: Evidence from Large Field Experiments on Alibaba, 2018.

Another Example

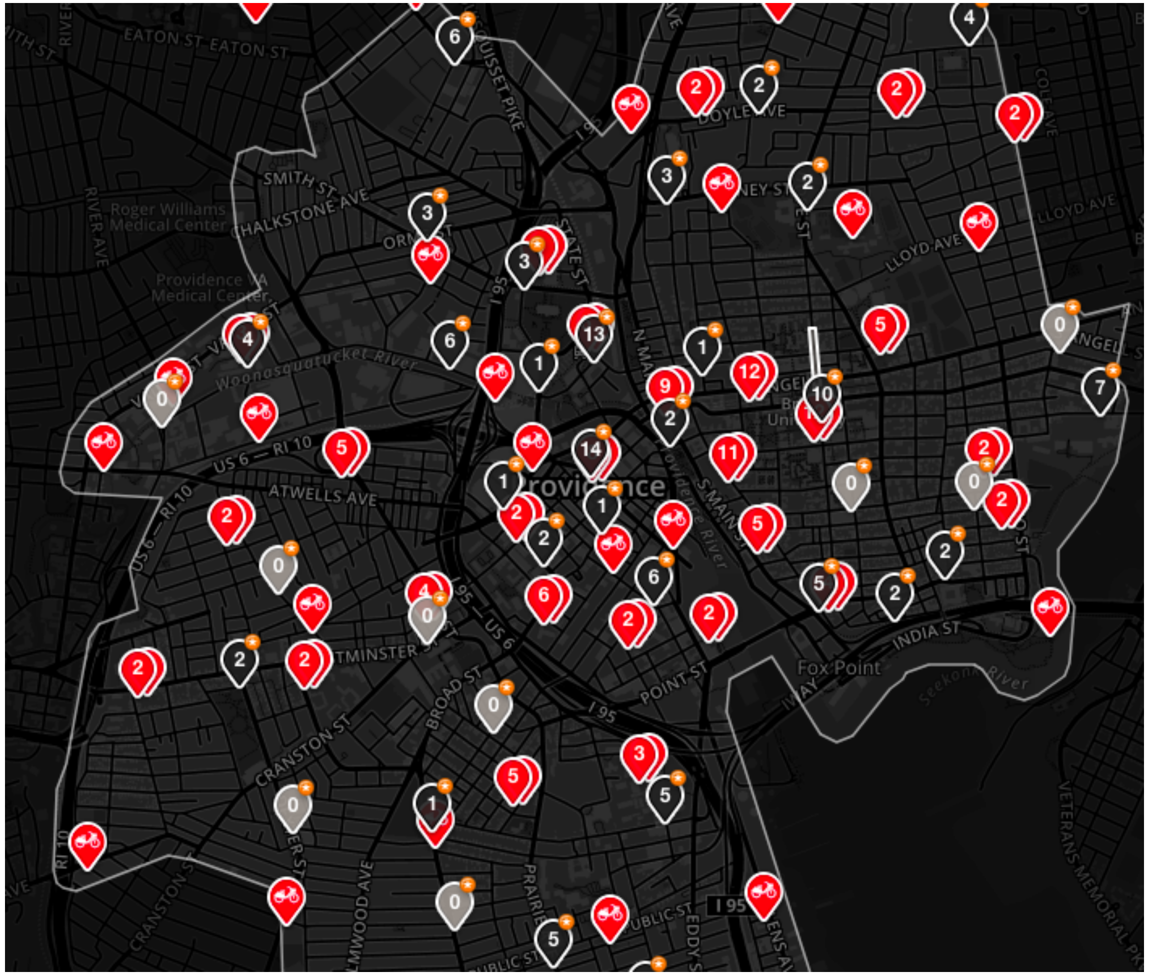
- Random utility model: $U_i = v_i + \epsilon_i$
- Customers buy the highest utility product offered (w/ utility > 0)
- Can be generalized as placing probabilities on possible preference lists

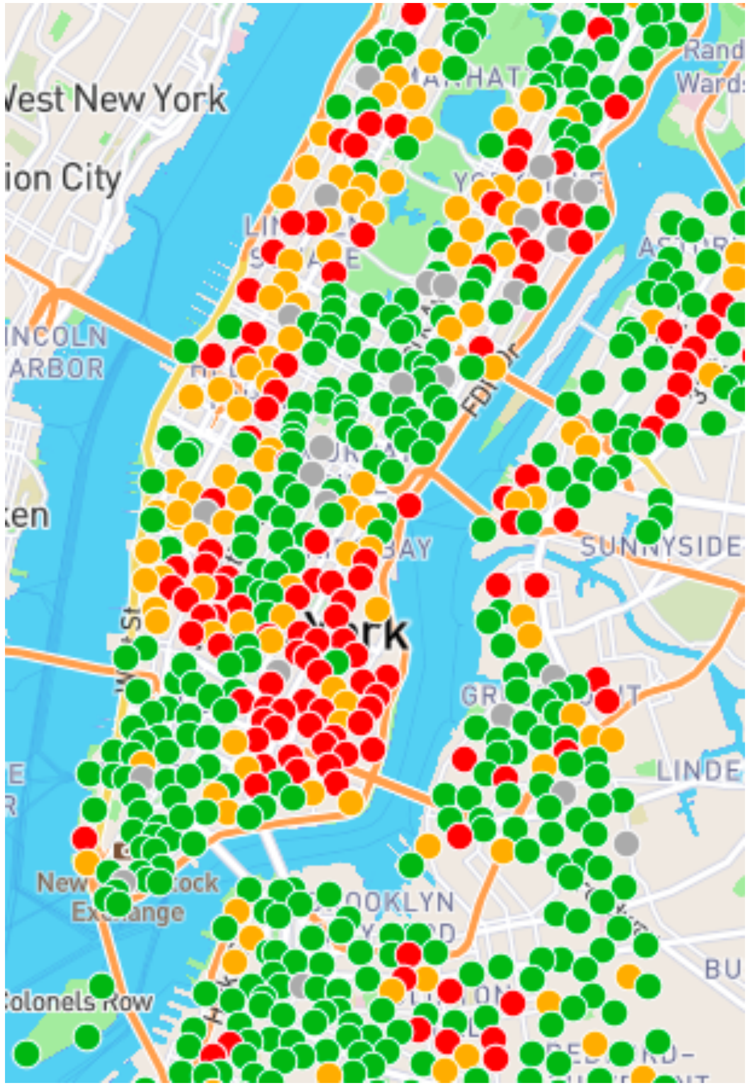


- Can find the most profitable assortments when lists are limited and there is evidence that customers don't consider many products in many settings.

Feldman, P., and Topaloglu, Technical Note: Assortment Optimization with Small Consideration Sets, Operations Research, 2018.

Bike-Share Systems





What Constitutes a *Good* Model for a Bike-Share System

- What do we mean by a model of the system?
- Interpretable model to accurately predict trip demand over the day – evidence that a simple model is most useful
- Ability to extend to dockless systems
- Accounting for correlation between stations
- Accounting for missing data
- Ability to optimize system parameters in model (where to place bikes, # bikes)

Decisions Made Based on the Model

- Number of bikes
- **Redistribution of bikes**
- Location and capacity of stations
- Boundary of system
- “Bike Angel” program
- **Repair of bikes**

Model informs objective function – minimize “unhappy” customers

Outline

- Examples of models
- Common problems arising
- Current interest in the problem
- Potential directions?

Example of a Simple Model (Singhvi et al.)

- Citi Bike group at Cornell used linear regression with simple covariates
- Clustering of stations improved the model – not surprising!
- $y_{\{i,j\}}$ = log(# bikes from station i to station j +1)
- Covariates: log taxi trips, log pick-up population, log drop-off population, indicator of borough pair, interactions between log taxi and indicators for borough pair
- Adj R^2 = 0.745

Example of a Simple Model (Singhvi et al.)

- Model fit for each time and weather category
- Clustering uses hard boundaries rather than thinking about a spatial distribution
- Model does not fit the data very well and there is no comparison to other approaches
- How could we simulate a day from this model? How could we count the number of unhappy customers?

Simulation of the Model

- Citi Bike group at Cornell works a lot with simulation to answer decision problems
- Two simulation-based methods – neither of which uses previous model!
  - Discrete-event simulation
  - Station independent queue model

Discrete-Event Simulation

- Using historical averages to find arrival rates for trips

$$\mu_{i,j,t} = (\#\ trips)/(time\ not\ empty)$$

- Each station has an independent Poisson process with rate

$$\mu_{i,t} = \sum \mu_{i,j,t}$$

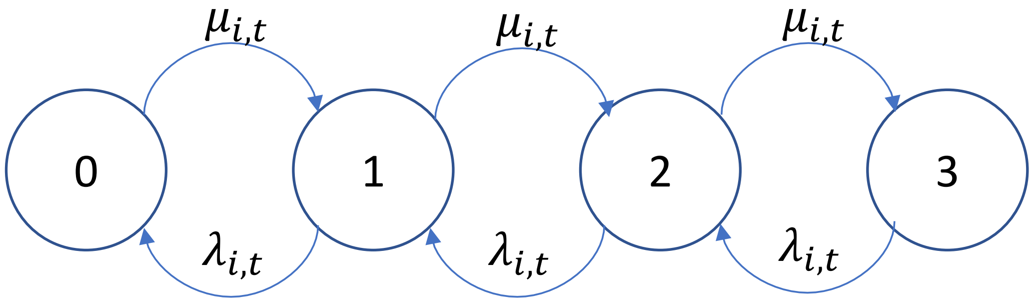
- If a bike is available, the user transitions to station j with probability

$$\frac{\mu_{i,j,t}}{\mu_{i,t}}$$

- Allowed to also look at 3 closest stations to j
- How could we extend this for a dockless system? Could we capture true demand better?

Station Independent Stock-Out Simulation

- For more complicated optimizations, sometimes assume independence between stations.
- Each station is treated as an M/M/1/k queue



- This model can be used as a warm start to an optimization of the whole system.

Added Elements

- Corrals
- Trikes
- Rebalancing trucks
- Bike Angels

Another Model (Li et al.)

- Hierarchical prediction model
- Cluster stations by traffic patterns and distance
- Predict total traffic in a time period using GBRT
- Assign this traffic to stations using past behavior
- Assign destinations using past behavior

Another Model (Li et al.)

- Reduced error compared to historical averaging and pure GBRT for check-in and check-out rates
- Not as continuous of a process, accumulation of error
- Clustering seems to improve model – don’t just cluster based on neighborhood

Other Models

- Poisson regression
- Negative binomial regression
- Neural networks

Current Interest and Dockless Systems

- Data is widely available
- More applications: scooters, electric bikes
- Dockless systems bring new challenges
- Throw deep learning at it!

Neural Network Approaches

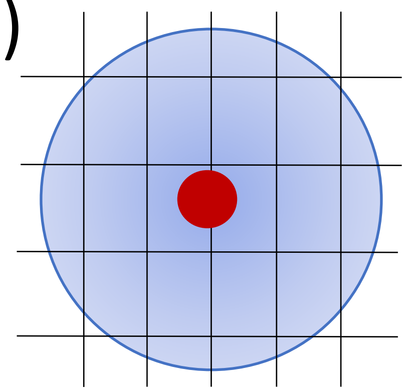
- M_t is distribution matrix at time t
- Covariates: $\{M_i, D_i, x_i \mid i = 1, \dots, t-1\}$
- D_i is distribution information (measure of uniformity) and x_i contains weather or time information
- Can capture spatial relationships through network
- How could we use this to estimate true demand? To optimize the system?

What Constitutes a *Good* Model for a Bike-Share System

- What do we mean by a model of the system?
- Interpretable model to accurately predict trip demand over the day – evidence that a simple model is most useful
- Ability to extend to dockless systems
- Accounting for correlation between stations
- Accounting for missing data
- Ability to optimize system parameters in model (where to place bikes, # bikes)

One Idea

- Modify historical averages to account for spatial information
- If a user picks up a bike, that demand could have come from a nearby location so weight by some likelihood (easiest would be uniform within some threshold t)



- $\mu_{i,t} = (weighted\ sum\ of\ trips)/(time\ with\ bike\ nearby)$
- Expect these new rates to be smooth and correlated.

Other Ideas?

- Spatio-temporal model of distribution
- Hierarchical model where we predict demand from distribution