

Problem Statement

The capstone project is with Rue Gilt Groupe building a **personal recommendation engine based on user navigational patterns**. Rue La La and Gilt are both e-commerce flash-sale websites focusing on fashion and home goods.

The project aims to **capture live user behavior as users browse on the site** and **recommend products based on the sequence of products user has browsed** to create a better, more seamless user experience. Thus, a Markov model was chosen to capture user-item interactions and provide personalized recommendations to users based on historical browsing behavior and the current product that user is viewing.

User browsing sequence

Product 1

Product 2

Product 3

Product 4

Navigation-based Recommendations

Model Overview

Historical Browsing Data

Clustering

Integer Optimization

Customized Solver

Evaluation

Parameter Tuning

Recommendation Optimization

I. Data Description and Data Processing

- Training: 9M observations (4/2019-7/2019); testing: 500K observations
- Transform historical user viewing data into product transitions

Transition Example

Item i	Item j	Transition Probability
Sweater	Skirt	23%
Dress	Coat	16%
Coat	Sweater	5%

II. Clustering Algorithm Design

User Transitions

Dress a

Dress b

Dress c

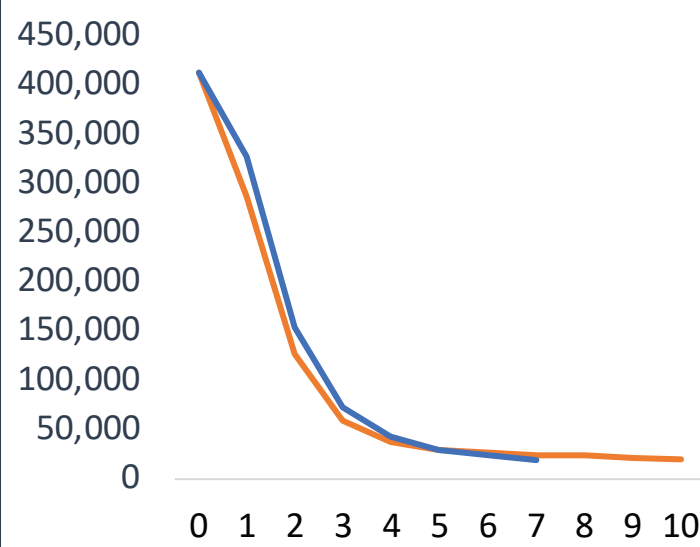
Group	$\sum \log P_{ij}$
1	-5.9
2	-20.4
3	-14.7
...	...

Based on historical site browsing data, cluster users:

- Randomly assign users to groups, calculate transition probability for all product transitions for a given user
 - Transition probability within group = $\frac{n_{ij}^g}{\sum_{i,x \in U^g} n_{ix}^g}$
- Adjust group assignment to maximize sum log of transition probabilities
- Re-calculate in-group transition probability
- Iterate until number of changes for entire epoch $\leq 5\%$ of users

i, j = product transitions observed for user u
 n_{ij}^g = number of product i to product j transitions in group g
 U^g = set of product to product transitions in group g
 $\sum_{i,x \in U^g} n_{ix}^g$ = number of observations in group g that transition out from product i

Number of Changes Per Epoch by Number of Groups



III. Recommendation Algorithm Design



Transition Probability

Probability of Exiting

Expected Length

40%

10%

40%

10%

20%

20%

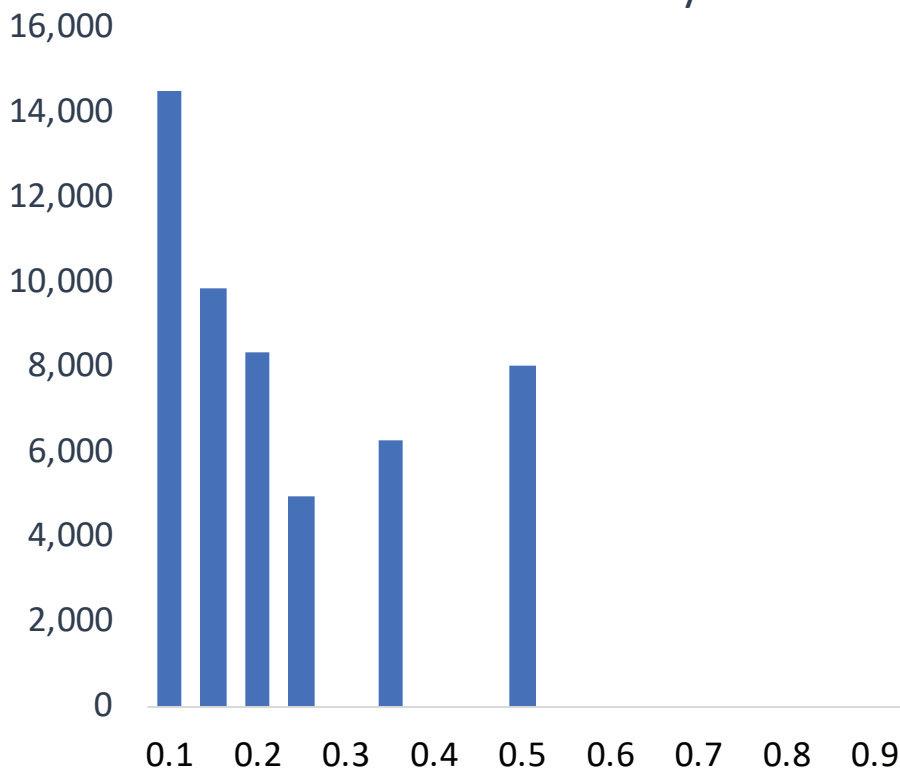
When a user logs on to Rue Gilt website, provide a list of product recommendations based on group assignment

- Rank recommendations based on a calculated score
 - $Score = \frac{p_{ij} - p_{j0} + E[l_j]}{3}$
 - p_{ij} = **Transition probability** from product i to product j (transition probability from current product to the recommended product)
 - p_{j0} = **Probability of ending** the browsing session after product j (recommended product)
 - $E[l_j]$ = **Expected length** for product j (recommended product)

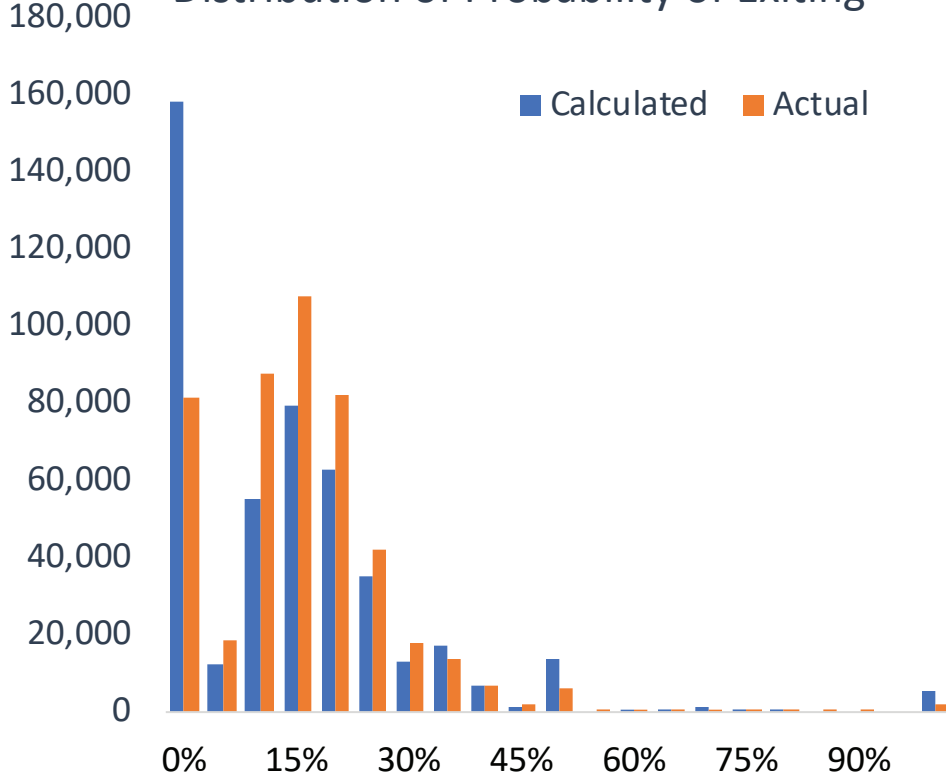
IV. Evaluation Mechanism Design

- Recommendation accuracy:** Percentage of hitting the next browsed product within the first 15 tries
- Reduce the probability of user immediately exiting the site**
- Increase expected length of browsing**

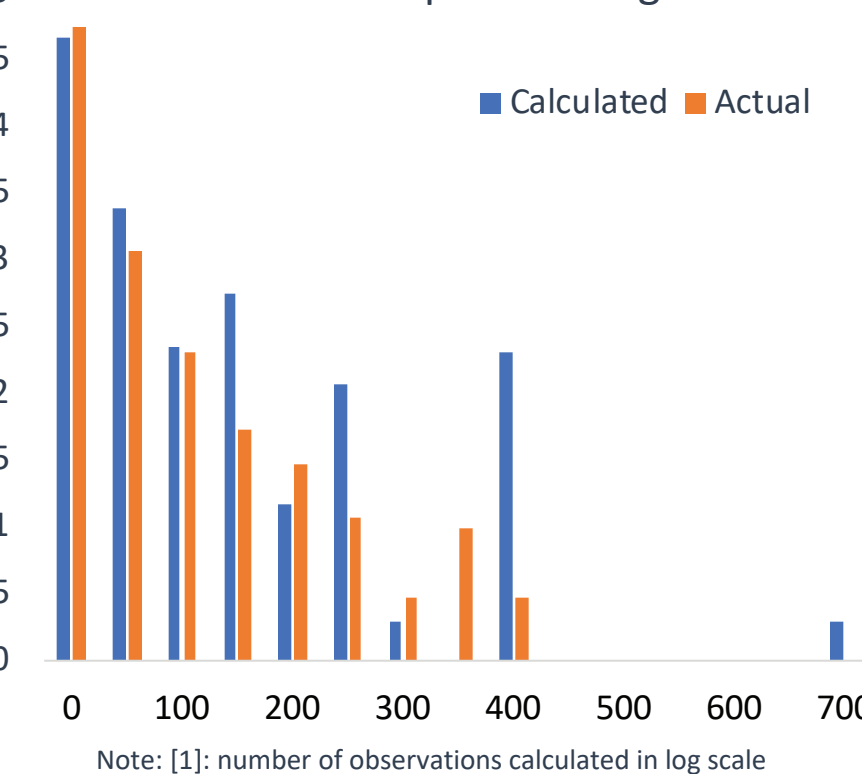
Distribution of Accuracy



Distribution of Probability of Exiting



Distribution of Expected Length ^[1]



Note: [1]: number of observations calculated in log scale

V. Business Impact

Evaluation Criteria	Expected Improvement
Percentage of success within the first 15 tries	12%
Probability of User Exiting	-10% per product transition
Expected increase in length of stay	+13 products/user

- Use user engagement as the primary business metric
- Generated by the extension of user length of stay on the website
- Assuming:
 - Each user will view **13 more products**
 - Historical data on order rates vs. number of products viewed, average revenue per order, and average daily visitors (with length > 1 filter)
 - 10-20%** customer adoption rate of new feature
- Business impact calculated as:
 - Expected increase in revenue = revenue with new feature - revenue without new feature = daily visitors * [order rate at new number of products viewed per user with new feature * adoption rate of new feature + order rate at current number of products viewed per user without new feature * (1 - adoption rate of new feature)] * average order value - daily visitors * order rate at current number of products viewed per user without new feature * average order value

1-3% expected annual revenue lift across Rue and Gilt