PERSONALIZED PROMOTIONS IN RETAIL SUPPLY CHAINS

SRIKANTH JAGABATHULA

Associate Professor of Tech, Ops, & Stats

Chief Scientist and Scientific Advisor

NYU STERN

arena

Walmart MAT Tech Council 2022
Broad overview of research

CANONICAL RETAILER
selling a subset of products from the market

DECISIONS
- Assortment: which brands to carry?
- Price: what prices or promotions to set?
- Inventory: how much of each product to carry?

PREDICTIONS
- demand for each product ~ function (assortment, prices)
ACCURATE DEMAND PREDICTIONS

E.g., $y = \beta^T x + \epsilon$

$y, x$

DATA

MODEL

METHOD

linear relationship between $x$ and $y$

Optimization method for OLS

- e.g., Newton’s method
- BFGS
- Gradient descent
- ....
Broad overview of research

**ACCURATE DEMAND PREDICTIONS**

- Noisy + Heterogeneous data
  - e.g., offline vs. online purchases
  - clicks vs. ratings vs. purchases
  - missing/censored data
  - sparse data for personalization
  - news/text data

- More complex choice models
  - e.g., rank-based
  - DAG-based
  - mixtures of Mallows
  - multichannel models
  - consideration set models

- Fast + accurate methods for large-scale data
  - e.g., MM algo. for nested logit models
  - nonparametric mixture distributions
  - nonparametric BLP for endogeneity
  - FW-method for rank-based models

- Our contributions to the ML literature
Personalizing Retail Promotions through a DAG-based Representation of Customer Preferences

Gustavo Vulcano
Professor of OM
School of Business
Torquato di Tella University
Argentina

Dmitry Mitrofanov
Asst. Prof. in Boston College
(former PhD student @ Stern)
Personalized promotions: an ongoing trend

- drive up sales
- increase both visits & basket size
- reduce competition
- stronger cust. relationship
- price discrimination

65% appreciate personalized prices
52% prefer weekly promotions

Forrester Consulting Study "Indiscriminate Promotions Cost Retailers" [2018]
Retailers want to move away from “mass promotions” to customized and targeted promotions.

Goal: CRAFT PROMOTION STRATEGIES USING HISTORICAL TRANSACTION DATA

Is the customer going to buy (@ full price) this item anyway?
- 52% of promotions go to customers who would pay full price
- 37% of customers are neutral or negative to such promotions

Which brand(s), if any, can we induce a switch to through promotions?
- Need to consider customer’s: brand loyalty, promotion sensitivity, brand sensitivity

How to optimize promotions, not independently, but jointly?
- Need to consider cannibalization and cross-product elasticities

Can save ~$60M

Forrester Consulting Study "Indiscriminate Promotions Cost Retailers" [2018]
Our focus: personalized demand predictions using panel data

Observations: for each customer and time period offer and promotion sets purchased product

offer set $S_t = \text{filled dots}$

promoted set $P_t = \text{green dots}$

purchase $i_t$

- $t = 1$
  - $S_1 = \{1, 2, 3, 4, 6, 7\}$, $P_1 = \{3, 6, 7\}$
  - $i_1 = 3$

- $t = 2$
  - $S_2 = \{1, 2, 3, 6, 7\}$, $P_2 = \{6, 7\}$
  - $i_2 = 3$

- $t = 3$
  - $S_3 = \{1, 2, 6, 7\}$, $P_3 = \{}$
  - $i_3 = 2$
Our focus: personalized demand predictions using panel data

- Historical purchase transactions tagged by customer ID
- Interpretable representations of customer preferences
- Customer 1
- Customer 2
- Optimal promotion strategy
Personalizing Retail Promotions through a DAG-based Representation of Customer Preferences

1. Model
   assumptions, data for estimation

2. Inference framework
   construction of DAGs, de-cycling, ML estimation

3. Numerical results
   results on real-world panel data from IRI dataset

4. Summary/Conclusions
   takeaway messages
1. Model assumptions, data for estimation

2. Inference framework construction of DAGs, de-cycling, ML estimation

3. Numerical results results on real-world panel data from IRI dataset

4. Summary/Conclusions takeaway messages
Choice model: preferences consistent across purchase instances

Partial order: DAG (Directed Acyclic Graph)

captures strong preferences of a customer
Choice model: preferences consistent across purchase instances

Partial order: DAG (Directed Acyclic Graph)

captures strong preferences of a customer

offer set

choice
Choice model: preferences consistent across purchase instances

Partial order: DAG (Directed Acyclic Graph)

captures strong preferences of a customer

offer set

choice

samples preference list consistent w/ DAG

- COLGATE
- CREST
- ORAL B
- SENSODYNE
- GERBER
- ARM & HAMMER
- NATURES GATE
- REMBRANDT [J&J]
- AQUAFRESH
- TOMS OF MAINE
IRI Academic Data Set

Weekly store sales and consumer panel data

31 product categories | 11 years
2001 - 2011

We analyzed year 2007 panel data

27 product categories | 83K user-category combinations | 1.2M transactions across 52 weeks
<table>
<thead>
<tr>
<th>Category name</th>
<th># Vendors</th>
<th># Customers retained</th>
<th>Avg. # trans.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beer</td>
<td>105</td>
<td>1154</td>
<td>7.11</td>
</tr>
<tr>
<td>Carbonated beverages</td>
<td>57</td>
<td>4387</td>
<td>17.63</td>
</tr>
<tr>
<td>Cigarettes</td>
<td>18</td>
<td>307</td>
<td>10.39</td>
</tr>
<tr>
<td>Coffee</td>
<td>73</td>
<td>2255</td>
<td>5.59</td>
</tr>
<tr>
<td>Cold cereal</td>
<td>45</td>
<td>3998</td>
<td>10.94</td>
</tr>
<tr>
<td>Deodorant</td>
<td>36</td>
<td>653</td>
<td>3.47</td>
</tr>
<tr>
<td>Facial tissue</td>
<td>13</td>
<td>2063</td>
<td>4.96</td>
</tr>
<tr>
<td>Frozen dinners/Entrees</td>
<td>97</td>
<td>3288</td>
<td>13.48</td>
</tr>
<tr>
<td>Frozen pizza</td>
<td>47</td>
<td>2846</td>
<td>7.83</td>
</tr>
<tr>
<td>Household cleaners</td>
<td>84</td>
<td>1699</td>
<td>4.14</td>
</tr>
<tr>
<td>Hot dogs</td>
<td>44</td>
<td>2187</td>
<td>3.82</td>
</tr>
<tr>
<td>Laundry detergent</td>
<td>24</td>
<td>2181</td>
<td>4.04</td>
</tr>
<tr>
<td>Margarine/Butter</td>
<td>19</td>
<td>2750</td>
<td>5.85</td>
</tr>
<tr>
<td>Mayonnaise</td>
<td>19</td>
<td>2386</td>
<td>3.28</td>
</tr>
<tr>
<td>Milk</td>
<td>37</td>
<td>4652</td>
<td>14.9</td>
</tr>
<tr>
<td>Mustard</td>
<td>60</td>
<td>2515</td>
<td>3.66</td>
</tr>
<tr>
<td>Paper towels</td>
<td>12</td>
<td>2051</td>
<td>5.2</td>
</tr>
<tr>
<td>Peanut butter</td>
<td>25</td>
<td>1923</td>
<td>3.89</td>
</tr>
<tr>
<td>Salt snacks</td>
<td>113</td>
<td>4446</td>
<td>15.09</td>
</tr>
<tr>
<td>Shampoo</td>
<td>81</td>
<td>738</td>
<td>3.73</td>
</tr>
<tr>
<td>Soup</td>
<td>123</td>
<td>4322</td>
<td>12.02</td>
</tr>
<tr>
<td>Spaghetti/Italian sauce</td>
<td>74</td>
<td>2698</td>
<td>5.46</td>
</tr>
<tr>
<td>Sugar substitutes</td>
<td>17</td>
<td>308</td>
<td>3.3</td>
</tr>
<tr>
<td>Toilet tissue</td>
<td>13</td>
<td>2817</td>
<td>5.1</td>
</tr>
<tr>
<td>Toothbrushes</td>
<td>52</td>
<td>499</td>
<td>3.06</td>
</tr>
<tr>
<td>Toothpaste</td>
<td>38</td>
<td>1186</td>
<td>3.58</td>
</tr>
<tr>
<td>Yogurt</td>
<td>32</td>
<td>3491</td>
<td>19.81</td>
</tr>
</tbody>
</table>
Training data summary

- aggregate items by vendors
- training: first 26 weeks
- retain customers ≥ 2 sales

After pre-processing

64K user-category combinations
600K transactions across 52 weeks

Average across 27 categories

50 vendors
2.3K customers
9.3 transactions per customer
Customized Individual Promotions: Model, Optimization, and Prediction

1. Model
   assumptions, data for estimation

2. Inference framework
   construction of DAGs, de-cycling, ML estimation

3. Numerical results
   results on real-world panel data from IRI dataset

4. Summary/Conclusions
   takeaway messages
DAG construction:
infer candidate edges from transactions

\[ t = 1; \text{offer set} = \{1, 2, 3, 4\}; \text{promotion set} = \{\} \]
DAG construction:
infer candidate edges from transactions
DAG construction:
infer candidate edges from transactions

candidate edges

edge (2, 3) may have been sampled

product 4 may not have been considered
DAG construction:
infer candidate edges from transactions
DAG construction: infer candidate edges from transactions

$t = 1; \text{offer set} = \{1, 2, 3, 4\}; \text{promotion set} = \{\}$

$purchased$

$t = 2; \text{offer set} = \{2, 5, 7\}; \text{promotion set} = \{\}$

$purchased$
DAG construction: infer candidate edges from transactions

$t = 2; \text{offer set} = \{2, 5, 7\}; \text{promotion set} = \{\}$

$t = 3; \text{offer set} = \{2, 5, 7\}; \text{promotion set} = \{\}$
DAG construction:
infer candidate edges from transactions

$t = 3; \text{offer set} = \{2, 5, 7\}; \text{promotion set} = \{}$

\[w_{5,2} = 2, \quad w_{5,7} = 2\]

$t = 4; \text{offer set} = \{4, 5\}; \text{promotion set} = \{}$

\[w_{5,2} = 2, \quad w_{5,7} = 2\]
DAG construction:
infer candidate edges from transactions

\[ w_{5,2} = 2 \quad w_{5,7} = 2 \]
De-cycling procedure removes spurious preference edges

spurious edges may be sampled from rank dist. from products not considered structural assumptions fewest sampled edges/largest DAG largest possible consideration sets while being consistent with the trans.

Find largest weighted sub-graph that is a DAG
To deal with promoted items, we maintain two copies of each item—promoted and nonpromoted.

$t = 1; \text{offer set} = \{1, 2, 3\}; \text{promotion set} = \{3\}$

purchased
<table>
<thead>
<tr>
<th>Category name</th>
<th># Vendors</th>
<th># Customers retained</th>
<th>% w/ Cycles</th>
<th>Size w/o cycles</th>
<th>Size w/ Cycles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beer</td>
<td>105</td>
<td>1154</td>
<td>60%</td>
<td>182.42</td>
<td>257.30</td>
</tr>
<tr>
<td>Carbonated beverages</td>
<td>57</td>
<td>4387</td>
<td>85%</td>
<td>83.61</td>
<td>130.70</td>
</tr>
<tr>
<td>Cigarettes</td>
<td>18</td>
<td>307</td>
<td>24%</td>
<td>28.62</td>
<td>36.85</td>
</tr>
<tr>
<td>Coffee</td>
<td>73</td>
<td>2255</td>
<td>51%</td>
<td>114.75</td>
<td>154.88</td>
</tr>
<tr>
<td>Cold cereal</td>
<td>45</td>
<td>3998</td>
<td>83%</td>
<td>76.91</td>
<td>132.91</td>
</tr>
<tr>
<td>Deodorant</td>
<td>36</td>
<td>653</td>
<td>47%</td>
<td>59.29</td>
<td>80.95</td>
</tr>
<tr>
<td>Facial tissue</td>
<td>13</td>
<td>2063</td>
<td>44%</td>
<td>19.92</td>
<td>26.00</td>
</tr>
<tr>
<td>Frozen dinners/Entrees</td>
<td>97</td>
<td>3288</td>
<td>71%</td>
<td>156.70</td>
<td>282.66</td>
</tr>
<tr>
<td>Frozen pizza</td>
<td>47</td>
<td>2946</td>
<td>59%</td>
<td>74.28</td>
<td>113.35</td>
</tr>
<tr>
<td>Household cleaners</td>
<td>84</td>
<td>1699</td>
<td>77%</td>
<td>140.72</td>
<td>199.94</td>
</tr>
<tr>
<td>Hot dogs</td>
<td>44</td>
<td>2187</td>
<td>46%</td>
<td>76.49</td>
<td>99.76</td>
</tr>
<tr>
<td>Laundry detergent</td>
<td>24</td>
<td>2181</td>
<td>37%</td>
<td>42.58</td>
<td>58.75</td>
</tr>
<tr>
<td>Margarine/Butter</td>
<td>19</td>
<td>2750</td>
<td>49%</td>
<td>39.39</td>
<td>58.00</td>
</tr>
<tr>
<td>Mayonnaise</td>
<td>19</td>
<td>2386</td>
<td>22%</td>
<td>30.70</td>
<td>37.30</td>
</tr>
<tr>
<td>Milk</td>
<td>37</td>
<td>4652</td>
<td>64%</td>
<td>69.38</td>
<td>102.10</td>
</tr>
<tr>
<td>Mustard</td>
<td>60</td>
<td>2515</td>
<td>64%</td>
<td>93.50</td>
<td>116.49</td>
</tr>
<tr>
<td>Paper towels</td>
<td>12</td>
<td>2051</td>
<td>52%</td>
<td>25.41</td>
<td>37.30</td>
</tr>
<tr>
<td>Peanut butter</td>
<td>25</td>
<td>1923</td>
<td>36%</td>
<td>37.59</td>
<td>47.25</td>
</tr>
<tr>
<td>Salt snacks</td>
<td>113</td>
<td>4446</td>
<td>86%</td>
<td>162.17</td>
<td>272.61</td>
</tr>
<tr>
<td>Shampoo</td>
<td>81</td>
<td>738</td>
<td>52%</td>
<td>113.97</td>
<td>147.62</td>
</tr>
<tr>
<td>Soup</td>
<td>123</td>
<td>4322</td>
<td>77%</td>
<td>186.19</td>
<td>291.87</td>
</tr>
<tr>
<td>Spaghetti/Italian sauce</td>
<td>74</td>
<td>2698</td>
<td>49%</td>
<td>108.28</td>
<td>149.98</td>
</tr>
<tr>
<td>Sugar substitutes</td>
<td>17</td>
<td>308</td>
<td>16%</td>
<td>24.09</td>
<td>28.22</td>
</tr>
<tr>
<td>Toilet tissue</td>
<td>13</td>
<td>2817</td>
<td>45%</td>
<td>27.75</td>
<td>41.76</td>
</tr>
<tr>
<td>Toothbrushes</td>
<td>52</td>
<td>499</td>
<td>48%</td>
<td>77.68</td>
<td>99.60</td>
</tr>
<tr>
<td>Toothpaste</td>
<td>38</td>
<td>1186</td>
<td>40%</td>
<td>57.14</td>
<td>74.26</td>
</tr>
<tr>
<td>Yogurt</td>
<td>32</td>
<td>3491</td>
<td>61%</td>
<td>49.91</td>
<td>77.22</td>
</tr>
</tbody>
</table>

% of customers with cycles and densities of DAGs
<table>
<thead>
<tr>
<th>Category name</th>
<th># Vendors</th>
<th># Customers retained</th>
<th>% w/ Cycles</th>
<th>Size w/o cycles</th>
<th>Size w/ Cycles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beer</td>
<td>105</td>
<td>1154</td>
<td>60%</td>
<td>182.42</td>
<td>257.30</td>
</tr>
<tr>
<td>Carbonated beverages</td>
<td>57</td>
<td>4387</td>
<td>85%</td>
<td>83.61</td>
<td>130.70</td>
</tr>
<tr>
<td>Cigarettes</td>
<td>18</td>
<td>307</td>
<td>24%</td>
<td>28.62</td>
<td>36.85</td>
</tr>
<tr>
<td>Coffee</td>
<td>73</td>
<td>2255</td>
<td>51%</td>
<td>114.75</td>
<td>154.88</td>
</tr>
<tr>
<td><strong>Cold cereal</strong></td>
<td>45</td>
<td>3998</td>
<td>83%</td>
<td>76.91</td>
<td>132.91</td>
</tr>
<tr>
<td>Deodorant</td>
<td>36</td>
<td>653</td>
<td>47%</td>
<td>59.29</td>
<td>80.95</td>
</tr>
<tr>
<td>Facial tissue</td>
<td>13</td>
<td>2063</td>
<td>44%</td>
<td>19.92</td>
<td>26.00</td>
</tr>
<tr>
<td>Frozen dinners/Entrees</td>
<td>97</td>
<td>3288</td>
<td>71%</td>
<td>156.70</td>
<td>282.66</td>
</tr>
<tr>
<td>Frozen pizza</td>
<td>47</td>
<td>2946</td>
<td>59%</td>
<td>74.28</td>
<td>113.35</td>
</tr>
<tr>
<td>Household cleaners</td>
<td>84</td>
<td>1699</td>
<td>77%</td>
<td>140.72</td>
<td>199.94</td>
</tr>
<tr>
<td>Hot dogs</td>
<td>44</td>
<td>2187</td>
<td>46%</td>
<td>76.49</td>
<td>99.76</td>
</tr>
<tr>
<td>Laundry detergent</td>
<td>24</td>
<td>2181</td>
<td>37%</td>
<td>42.58</td>
<td>58.75</td>
</tr>
<tr>
<td>Margarine/Butter</td>
<td>19</td>
<td>2750</td>
<td>49%</td>
<td>39.39</td>
<td>58.00</td>
</tr>
<tr>
<td>Mayonnaise</td>
<td>19</td>
<td>2386</td>
<td>22%</td>
<td>30.70</td>
<td>37.30</td>
</tr>
<tr>
<td>Milk</td>
<td>37</td>
<td>4652</td>
<td>64%</td>
<td>69.38</td>
<td>102.10</td>
</tr>
<tr>
<td>Mustard</td>
<td>60</td>
<td>2515</td>
<td>64%</td>
<td>93.50</td>
<td>116.49</td>
</tr>
<tr>
<td>Paper towels</td>
<td>13</td>
<td>2051</td>
<td>52%</td>
<td>35.41</td>
<td>37.30</td>
</tr>
</tbody>
</table>

% of customers with cycles and densities of DAGs

54%

customers with cycles

avg. # of edges in DAGs

73 in cust. 150 in cust.

w/o cycles w/ cycles

avg. # of trans. per cust. 16 avg. offer set size
- brand loyal to “Philip Morris”, then “Reynolds Light”
- might switch to “Howard”, “Reynolds Light”, or “Commonwealth Light” if promoted and “Philip Morris” is stocked out
1. Model assumptions, data for estimation

2. Inference framework construction of DAGs, de-cycling, ML estimation

3. Numerical results results on real-world panel data from IRI dataset

4. Summary/Conclusions takeaway messages
Experiments conducted: one step-ahead prediction

\[ U = \text{user set} \quad N = \text{product set} \quad T = \# \text{ of discrete time periods} \]

For any \( t = 1, 2, 3, \ldots, T \)

**Given**
- everything until time period \( t \)
- (offer sets, promoted items, purchases of users in \( U \))
- \( S_{t+1} = \text{offer set in period} \ t+1 \)
- \( P_{t+1} = \text{promoted items in period} \ t+1 \)
- \( U_{t+1} = \text{users purchasing in period} \ t+1 \)

**Prediction**
- \( f_u(i, t+1) = 1 \) if \( i \) has highest choice probability for \( u \) in period \( t+1 \)
  - for all \( u \in U_{t+1}, \ i \in S_{t+1} \)

\[
X^2 \text{ score} = \frac{1}{|U||N|} \sum_{u \in U, i \in N} \frac{(n_{ui} - \hat{n}_{ui})^2}{0.5 + \hat{n}_{ui}} \quad \hat{n}_{ui} = \sum_{t \in T} f_u(i, t) \quad \text{similar to } \chi^2 \text{ score } \frac{(O - E)^2}{E} \\
\text{miss rate} = \frac{1}{|U||T|} \sum_{u \in U, t \in T} I[f_u(a_{j_u,t}) \neq 1] \quad \# \text{ obs. purchase in } t \text{ of } u \quad \text{lower is better}
\]
Lower is better
Key takeaways:

- Even single-class MNL outperforms RPL for most categories
- RPL has more parameters. Time: our method ~ 10 secs, RPL ~ 67 mins
- Gains higher for cust. w/o cycles because of strong prefs.
- De-cycling extends coverage to all customers
- single-class: heterogeneity through DAGs
- multi-class: additional heterogeneity through classes
The DAG already provides insights on which products to promote for a given offer set

Decision: which items to put on promotion?

offer set = \{1, 2, 3, 4\}

- Item 3 will NOT be purchased whether promoted or not.
- Item 1 will be purchased only when promoted, if item 2 is not promoted.
revenue from existing promotion strategy

38

revenue from optimized promotion strategy

predicted revenue on hold-out offer sets

observed revenue on hold-out offer sets

24%

revenue gain from personalizing promotions

HUGE revenue opportunity

impr w./o. mass = 23.93%

impr w. mass = 16.61%

MAE = 6.34%

accurate revenue predictions from our method
Personalizing Retail Promotions through a DAG-based Representation of Customer Preferences

1. Model assumptions, data for estimation
2. Inference framework construction of DAGs, de-cycling, ML estimation
3. Numerical results results on real-world panel data from IRI dataset
4. Summary/Conclusions takeaway messages
Summary and key findings

**KEY CONTRIBUTION**
Methodology to design personalized promotions from panel data through DAGs

**MAIN TAKEAWAYS**
- DAGs: provide rich representation of preferences
- Better predictive accuracy: more than 15% better predictive accuracy compared to state-of-the-art benchmarks
- HUGE revenue opportunity: up to 24% revenue gain w/ personalized promotions
Active Learning for Personalized Promotions in Supply Chains
B2B promotions from a brand to a store require an active learning strategy

PROBLEM
optimize SKU promotions to maximize profit

CHALLENGES
- rapid demand shifts (seasonality, SKU intros)
- insufficient historical price variation
- time and cross-SKU cannibalization effects

Our Solution:
ACTIVE LEARNING USING ADVERSARIAL BANDIT FRAMEWORK