Online Retailing is Growing Fast

Digital revolution – All about the data!

US consumer online retail spending will grow by 55% in the next five years.

- Forrester Research, 2018
Motivation for Personalization

• You can shop at **any time** and **from anywhere**

• Customers visit Internet stores **more frequently**

• **More availability of data**
  – User logins and cookies
  – Virtual shopping cart
  – Items put in shopping cart (“bookmark” items of interest), wish lists

• Can infer information (features) on customers

---

Online Retail Marketplace -- Personalization

• **“54% of customers** who already receive customized messages indicated they were **more likely to buy from vendors who interacted** with them via targeted marketing.”

• “Many shoppers—**59%**—also **had positive feelings** about **receiving custom offers from the places where they shop**”

**Create offerings that are better tailored to a customer’s needs.**

**Educate consumers about products they may not otherwise be aware of.**

**Improve customer experience.**
Who Can Set Trends?

Advertisement - Celebrities

Advertisement – Fashion Bloggers
But any of us can create trends too!
Apart from celebrities and fashion bloggers, others can start underlying trends? Examples!

Benefits of Identifying the Influencers

Can we detect “somehow” underlying trends between (groups of) consumers and improve demand forecast?

How can we leverage the underlying trends to improve targeted offers?
Social Data Can Help

But what if this data is not available?

What are other factors that can help?

Utilize to predict demand at a personalized level?

Targeted Offers

Sell the right product

At the right time

To the right customer

At the right price

Obeying business rules
We Bring Together Two Important Aspects

Outline
- The Problem (what and why)
- The Data
- The Approach (Model)
  - Personalized Demand Estimation
  - Targeted Offer Optimization
- Impact from Data
Some Relevant Literature

- Cohen, Harsha. 2015
- Candogan, Bimpikis, and Ozdaglar. Operations Research 2012
- Blattberg, Nolin. '90
- Talluri, van Ruzin '05
- Cohen, Leung, Panchamgam, Perakis, Smith '16
- Bai, Kim, Cohen, Panchamgam, Perakis, Segev '16
- Cohen, Harsha. 2015
- Candogan, Bimpikis, and Ozdaglar. Operations Research 2012
- Blattberg, Nolin. '90
- Talluri, van Ruzin '05
- Cohen, Leung, Panchamgam, Perakis, Smith '16
- Bai, Kim, Cohen, Panchamgam, Perakis, Segev '16
- Aral and Walker. Management Science 2011
- Girorezael, Nazerzadeh, Rusin. '14
- Johnson, Lee, Simchi-Levi '15
- Gallego, Li, Triung, Wang '16
- Chen, Owen, Pixton, Simchi-Levi '16

Contributions

Study a model for personalized demand estimation with customer to customer trends and targeted promotion optimization.
Overall Goal

• Target the *right customers* with promotions so that they purchase and potentially create trends so that others purchase too

• Two steps:
  • First understand **purchase probability** (demand)
  • Then determine **promotion targeting strategies**
Methodology: Base Demand and Trend Model

**Base Demand** for customer (group of customers) A to buy the item
Estimation through a demand model with customer features

\[
\text{Demand} = \text{Base demand} + \text{trends}
\]

How to detect and include customer to customer trends?

Detecting Customer to Customer Trends

During the selling season customers come in and purchase

Possible trends between (groups) of customers

Granger causality: customers can only "influence" / create trends to other customers who buy later
**Transaction Data**

Transaction data: frequency of purchase patterns to find customer to customer trends

**Trend Model from Transaction Data**

Represent the purchase patterns as graphs  
Even when underlying network structure is not known
Interested in estimating the **underlying customer to customer trend**

Effect by which a (group of) customer(s) (say A) “changes” another (group of) customer’s purchase probability (say C)

---

**Base Model**

\[
\begin{align*}
b_{A1} &= q_{A1} + \epsilon_{A1} \\
b_{C1} &= q_{C1} + \epsilon_{C1}
\end{align*}
\]

**Base purchase probability** \(q_{Ci} \) of customer \(c=A\) for item \(i\) at time \(t=1\)

**Purchase probability** \(b_{ci} \) that customer \(c=A\) buys item \(i\) at time \(t=1\)

- Build base model from transaction data
Adding Customer Trend

\[ b_{A11} = q_{A11} + \epsilon_{A11} \]
\[ b_{C12} = q_{C12} + p_{AC} + \epsilon_{C12} \]

Trend probability \( p_{c'c} \) from customer \( c' (=A) \) to customer \( c (=C) \)

- Obtain graphs from past transactions
- With enough items/data obtain the true customer to customer trend graph

Formulating the Customer-to Customer Trend Model

- General form of the customer trend model

\[ b_{cty} = p_{cty} + \sum_{t=1}^{\ell-1} \sum_{c' \in M} p_{c'ct} b_{cty'} + \epsilon_{cty} \]

- Solve via MLE but problem nonconvex.
- Approximate solution approach:
  1. Estimate \textbf{base purchase probability} with standard demand model (logistic)
  2. Remove base purchase probability to get \textbf{standardized purchase probability}
  3. Estimate \textbf{customer-to-customer trend probability}
Estimation Step 1

1. Estimate \textbf{base purchase probability} with standard demand model

   Start by thinking as if trend does not exist

   - Impose a structure on \( q_{cit} \), for example logistic regression

   \[
   q_{cit} = \frac{\exp(x_{cit}^T \beta)}{1 + \exp(x_{cit}^T \beta)}
   \]

   \( x_{cit}^T \): features associated with the particular transaction

   (e.g. \textit{seasonality, brand, style, promotion} among others)

   - Fitting the logistic regression yields parameters \( \hat{\beta} \) (MLE)

Estimation Step 2

2. Compute \textbf{empirical probabilities} \( b_{cit} \)

   - Remove base purchase probability to get \textbf{standardized purchase probability}

   - Standardize/Deseasonalize the purchase probability

   - Predict the base purchase probability for each \((c, i, t)\)

   \[
   \hat{q}_{cit} = \frac{\exp(x_{cit}^T \hat{\beta})}{1 + \exp(x_{cit}^T \hat{\beta})}
   \]

   - \textbf{Standardize} by subtracting predicted base purchase probability

   \[
   b_{cit}^s = b_{cit} - \hat{q}_{cit}
   \]
Estimation Step 3

3. Estimate **customer to customer trend probability** by minimizing error
   – Describe standardized/deseasonalized data with the customer to customer trend model

   - Solve the optimization model (bounded variables least squares)
     - \( \min \sum_c \sum_i \sum_t \epsilon_{cit}^2 + \lambda \sum_{(c',c)} p_{c'c} \)
     - \( \epsilon_{cit} = b^S_{cit} - \sum_{c' \neq c} \sum_{t' = t-M}^{t-1} p_{c'c} b_{c'it'} \quad \forall c, i, t \)
     - \( 0 \leq p_{c'c} \leq 1 \quad \forall c', c \)
   
   Resulting in **estimated customer to customer trend probabilities** \( \hat{p}_{c'c} \)
   
   **Bounded variables least squared with Lasso**

---

**Endogeneity**

- Consider linear regression model: \( y = X\beta + \epsilon \)

- Estimation with **ordinary least squares** leads to \( \hat{\beta} = (X^TX)^{-1}X^Ty \)

- Assuming \( E[\epsilon|X] = 0 \), we have **consistency**
  \( \hat{\beta} \to E[\hat{\beta}] = E[(X^TX)^{-1}X^T(\beta + \epsilon)] = \beta + E[(X^TX)^{-1}X^TE[\epsilon|X]] = \beta \)

- But often \( x \) can be **endogenous** so that \( E[\epsilon|X] \neq 0 \)
Instrumental Variables

• Propose **instrumental variable** $Z$ for **endogenous variable** $X$ so that
  1. $Z$ is uncorrelated with the error term $\varepsilon$: $E[Z^T \varepsilon] = 0$
  2. $Z$ is correlated with the endogenous variable $X$: $E[Z^TX] \neq 0$

• Estimation with **two-stage least squares** leads to $\hat{\beta}^{IV} = (Z^TX)^{-1}Z^Ty$
  1. Fit OLS of $X$ on $Z$ to estimate $\hat{\mu}$ and predict $\hat{X} = \hat{\mu}^TZ$
  2. Fit OLS of $y$ on $\hat{X}$ to estimate $\hat{\beta}^{IV}$

• Under the conditions of an instrumental variable, we have **consistency**
  $$\hat{\beta}^{IV} = (Z^TX)^{-1}Z^T(X\beta + \epsilon) = \beta + (Z^TX)^{-1}Z^T\epsilon \to \beta$$

Context of Customer Trends

• For each $c$ separately, we fit a linear regression
  $$b_{cit} - \hat{q}_{cit} = \sum_{c'=1}^{n} p_{c'c} \sum_{t'=t-M}^{t-1} b_{c'it'} + \epsilon_{cit}$$

• Comparing with the linear regression model
  $$y = \beta^T X + \epsilon$$

• Need one instrument for every column of $X$ or for every $\sum_{t'=t-M}^{t-1} b_{c'it'}$
Instrumental Variables for Customer Trends

\[ y = \beta^T X + \epsilon \]

- For each \( c \) separately, we fit a linear regression

\[ b_{cit} - \hat{q}_{cit} = \sum_{c' = 1}^{n} \rho_{c'c} \sum_{t' = t-M}^{t-1} b_{c',it'} + \epsilon_{cit} \]

\[ \text{Purchases during M+1 periods ago} \]

- Propose \( b_{c,i,t-M-1} \) (for all \( \tilde{c} \)) as instrument, needs to satisfy:
  1. \( b_{c,i,t-M-1} \) are uncorrelated with \( \epsilon_{cit} \)
  2. \( b_{c,i,t-M-1} \) are correlated with \( \sum_{t' = t-M}^{t-1} b_{c',it'} \)

\[ \text{Purchases far in the past are not affecting current purchases} \]

\[ \text{Purchases not too far in the past are affecting recent purchases} \]

Estimating Customer Trends

\[ y = \beta^T X + \epsilon \]

- For each \( c \) separately, we fit a linear regression

\[ b_{cit} - \hat{q}_{cit} = \sum_{c' = 1}^{n} \rho_{c'c} \sum_{t' = t-M}^{t-1} b_{c',it'} + \epsilon_{cit} \]

\[ \text{Purchases during M+1 periods ago} \]

- Estimation procedure involves the same two-stage approach
  1. Fit OLS of \( \sum_{t' = t-M}^{t-1} b_{c',it'} \) on \( b_{c,i,t-M-1} \) (for all \( \tilde{c} \)) to estimate \( \hat{\gamma} \)

\[ \frac{X}{\sum_{t' = t-M}^{t-1} b_{c',it'}} = \sum_{c = 1}^{n} \hat{\gamma}_c b_{c,i,t-M-1} \]

\[ \text{Condition 2: F tests p-values < 2e-16} \]
\[ \text{Craig-Donald test > 46.62} \]

2. Fit bounded variables OLS of \( b_{cit} - \hat{q}_{cit} \) on \( \sum_{t' = t-M}^{t-1} b_{c',it'} \) (for all \( c' \)) to estimate \( \hat{\rho}_{c',c} \)
Convergence

**Theorem:** Under the following two assumptions

1. Estimation of $q_{cktr}$ is consistent: $q_{cktr} \rightarrow q_{cktr}$ as $T$ or $K \rightarrow \infty$
2. Trend only occurs one time period ahead

Then, the customer-to-customer trend estimation algorithm estimates $p_{c'c}$ consistently

$$\hat{p}_{c'c} \rightarrow p_{c'c} \text{ (as } k \text{ or } T \rightarrow \infty)$$

Fashion Data

- Tier 1 Fashion Retailer ~600 stores, Rev: 1-5B
- **Transaction data features**
  - Customer: Loyalty program ID, zipcode
  - Item: Women’s dressy tops category
  - Time: January 2014 – June 2016
  - Store: US, city, state, square footage
  - Price: Regular and promo prices
- **Size per state**
  - ~10k customers
  - ~4k items
  - ~20 stores
In Summary: Aggregation

- **Data aggregation**
  - Customers: Over location and spending
  - Item: Over styles (400 most sold)
  - Time: Over weeks (159 weeks)

- **Size per state**
  - ~60 customer types
  - ~150 customers/customer type
  - ~400 styles
  - ~10 items/style

Prediction Metrics

- Fit for actuals $y_i$ and predictions $\hat{y}_i$ ($i = 1, \ldots, n$):
  
  \[
  MAPE(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{\hat{y}_i - y_i}{y_i} \right|
  \]

  \[
  WMAPE(y, \hat{y}) = \sum_{i=1}^{n} \left| \frac{\hat{y}_i - y_i}{y_i} \right| \frac{y_i}{\sum_{i=1}^{n} y_i}
  \]

- We use **Weighted Mean Absolute Percentage Error** (WMAPE)
  - Looks at percentage difference between prediction $\hat{y}_i$ and actual $y_i$
  - Take **weighted average** of these percentage differences
  - **Higher sales** observations are **weighted higher**
Demand Prediction Results

Relative improvement of WMAPE by including customer to customer trend

Comparison of out-of-sample WMAPE of the base demand model and the customer trend demand model for different departments in the states of Ohio and Utah

<table>
<thead>
<tr>
<th>Department State</th>
<th>Base</th>
<th>Customer Trend</th>
<th>Relative Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mens Knit Tops Ohio</td>
<td>0.5213</td>
<td>0.4683</td>
<td>10.2%</td>
</tr>
<tr>
<td>Utah</td>
<td>0.5024</td>
<td>0.4845</td>
<td>3.6%</td>
</tr>
<tr>
<td>Women’s Dressy Ohio</td>
<td>0.4548</td>
<td>0.3951</td>
<td>13.1%</td>
</tr>
<tr>
<td>Tops Utah</td>
<td>0.4322</td>
<td>0.3636</td>
<td>15.9%</td>
</tr>
</tbody>
</table>

How WMAPE changes

[Graph showing changes in WMAPE with changes in memory M]
Results in Ohio

• Stores in two seemingly distinct areas of Ohio
• Areas are integrated
  – Northern center Cleveland
  – Southern center Dayton
• Cleveland in the north and Dayton in the south link the areas

Promotion Targeting Problem

So far we’ve estimated demand with customer behavior and customer to customer trends

Target promotions to maximize expected profit
Promotion Targeting Problem

- Target the right customers with the right promotions, for the right items and at the right time.
- Goal is to maximize overall profit (say over quarter)

- Business constraints:
  - Do promotions only at specific time periods
  - Limit the total number of promotions
  - No touch constraints: two successive promotions are separated by at least some weeks

Promotion Targeting Model

\[
\begin{align*}
\text{Expected revenue} & \quad \text{Backorder shipping cost} \\
\max & \quad \sum_{c=1}^{n} \sum_{t=1}^{T} \sum_{i=1}^{m} r_{cit}(N_c b_{cit}) - \sum_{i=1}^{m} \sum_{l=1}^{L} \left( \sum_{c \in C_l} \sum_{t=1}^{T} (N_c b_{cit}) - l_{il} \right) \\
\text{s.t.} & \quad b_{cit} = q_{cit} \Phi(r_{cit}) + \sum_{c'=1}^{n} \sum_{t'=t-M}^{t-1} p_{c'} c_{c't'} \quad \forall c, i, t \\
& \quad r_{cit} = d_i (1 - y_{cit}) + d_i y_{cit} \quad \forall c, i, t \\
& \quad \sum_{c=1}^{n} \sum_{i=1}^{m} \sum_{t=1}^{T} y_{cit} \leq L \\
& \quad y_{cit} \in \{0,1\} \quad \forall c, i, t
\end{align*}
\]

The problem is **integer** and **nonlinear**

The problem is **NP-hard**

Reduction from the set cover problem

- Define price decision variable
- Limit on the number of promotions
- Decision variable indicating whether a promotion is offered to customer $c$ for item $i$ at time $t$
Fast Algorithm for Promotion Targeting

**Adaptive Greedy Algorithm:**
- Calculate iteratively the potential increase in revenue from promoting each (group of) customer(s)
- Select the customer(s) (group of) with the highest potential for promotion
- Do up to \( L \) Promos

<table>
<thead>
<tr>
<th>Potential</th>
<th>Limit of ( L = 2 ) Promotions</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

When is the Promotion Targeting Problem Submodular?

**Theorem:** Under the following setting,
1. The base purchase probability function is a convex function of the promotion policy
2. The items are complementary or independent

The Promotion Targeting Problem is Submodular

The Adaptive Greedy finds

\[
(1 - 1/e)f(S^{opt})
\]

Approximation ratio
C-Submodular Functions

For every two promotion policies $\Phi, \bar{\Phi}$, such that every costumer that $r_{cit}^{\Phi} \leq r_{cit}^{\bar{\Phi}}$ and given a customer $c$ and item $i$ at a time period $t$ for which $r_{cit}^{\Phi} = r_{cit}^{\bar{\Phi}} = d_0$,

$$R(\Phi, r_{cit}^{\Phi} = d_K) - R(\bar{\Phi}) + C \geq R(\Phi, r_{cit}^{\Phi} = d_K) - R(\Phi)$$

Adaptive Greedy Algorithm and C-Submodular Functions

**Theorem:** For problems with C-submodular objective, the Greedy algorithm finds

$$\left(1 - \frac{1}{e}\right)f(S^{opt}) - LC \left(1 - \frac{1}{e \cdot (L - 1)}\right)$$

Approximation ratio.
## Summery of Approximation Ratio - Adaptive Greedy

<table>
<thead>
<tr>
<th>Description</th>
<th>Approximation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Submodular</td>
<td>$\left(1 - \frac{1}{e}\right) f(S^{\text{opt}})$</td>
</tr>
<tr>
<td>C-Submodular</td>
<td>$\left(1 - \frac{1}{e}\right) f(S^{\text{opt}}) - LC \left(1 - \frac{1}{e \cdot (L - 1)}\right)$</td>
</tr>
</tbody>
</table>

### Profit Impact in Fashion: Ohio

**Current promotion policy**

- Cincinnati
- Dayton
- Columbus
- Toledo
- Youngstown

**Optimal promotion policy**

- Profit improvement of **3.7%-10.6%**
Conclusions

Demand Model w. Customer to Customer Trend
- Interpretable model to estimate trends among customers
- Uncovers customer relationships when there is no clear underlying connection and consistently estimates trend

Optimizing Promotion Targeting
- Model to optimize the targeting of promotions
- Adaptive Greedy algorithm to solve problem
- Adaptive Greedy algorithm good approximation under more general setting

Results on Fashion Data
- Generate insights on fashion data from Oracle client
- Including customer trend improves WMAPE by 3-15%
- Promotion targeting improves profits by 3-10%

Introduced and analyzed a model for Personalized demand estimation with customer to customer trend AND targeted promotion optimization.

Thank You!

Questions?