SHOPPER: A Probabilistic Model of Consumer Choice with Substitutes and Complements

Francisco J. R. Ruiz

With Susan Athey and David M. Blei

June 14th, 2018

Introduction

Large-scale market basket data

Why Analyzing Market Basket Data?

- Understand consumer behavior
- Make predictions about demand
- Predict response to promotions or price changes
- Form personalized recommendations

Market Basket Data is Complex

Challenges:

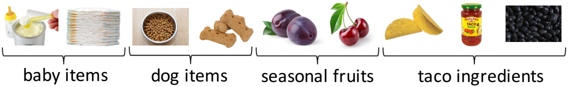
- Many interrelated forces at play
- Some are unobserved

Market Basket Data is Complex

Challenges:

- Many interrelated forces at play
- Some are unobserved

Example:



Our Contribution

SHOPPER:

- A sequential probabilistic model of shopping baskets
- Interpretable components
- Captures user heterogeneity, seasonal effects, prices
- Forms predictions under price changes

Our Contribution

SHOPPER:

- An efficient posterior inference algorithm
- Empirical study:
  - Accurate predictions under price interventions
  - Helps identify complements and substitutes

Model: Items Are Chosen Sequentially

- Customer walks into the store and chooses item sequentially
- At each step, chooses over items that are not in the basket
- The sequence ends when customer chooses the “checkout item”

Model: Unobserved Item/User Attributes

- The probability of choosing an item depends on latent factors
- Item attributes: $\alpha_c \in \mathbb{R}^K$
- User preferences: $\theta_u \in \mathbb{R}^K$
- The inner product $\theta_u^\top \alpha_c$ determines the probability

SHOPPER: Vanilla Version

- Item interaction coefficients: $\rho_c \in \mathbb{R}^K$
- Define a utility for each item c at each step i in trip t
- The (mean) utility depends on previously chosen items,

$$\Psi(c, \underbrace{\mathbf{y}_{t,i-1}}_{\substack{\text{items in}\\ \text{basket}}}) = \underbrace{\psi_{tc}}_{\substack{\text{user preferences:}\\ \theta_u^\top \alpha_c}} + \underbrace{\rho_c^\top \left(\frac{1}{i-1} \sum_{j=1}^{i-1} \alpha_{y_{tj}} \right)}_{\text{item interactions}}$$

- In terms of probabilities,

$$p(y_{ti} = c \mid \mathbf{y}_{t,i-1}) = \frac{\exp\{\Psi(c, \mathbf{y}_{t,i-1})\}}{\sum_{c' \notin \mathbf{y}_{t,i-1}} \exp\{\Psi(c', \mathbf{y}_{t,i-1})\}}$$

Baskets as Unordered Set of Items

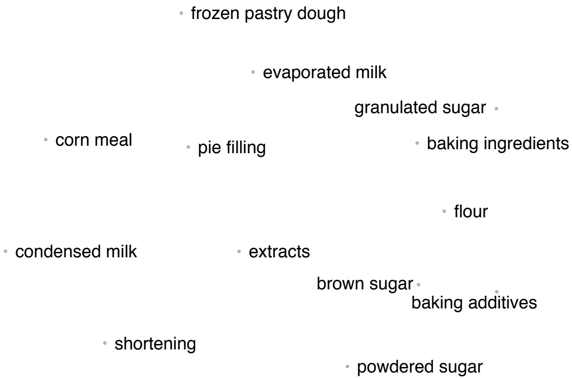
- Probability of an ordered basket (product of individual choices),

$$p(\mathbf{y}_t \mid \rho, \alpha, \theta) = \prod_{i=1}^{n_t} p(y_{ti} \mid \mathbf{y}_{t,i-1}, \rho, \alpha, \theta)$$

- When the ordering is not observed, sum over all possible orderings

$$p(\mathcal{Y}_t \mid \rho, \alpha, \theta) = \sum_{\pi} p(\mathbf{y}_{t,\pi} \mid \rho, \alpha, \theta)$$

Item Attributes Capture Meaningful Representations



(Zoom on 2D projection of item space α_c)

Price and Seasonal Effects are Additive Components

$$\Psi(c, \underbrace{\mathbf{y}_{t,i-1}}_{\substack{\text{items in} \\ \text{basket}}}) = \underbrace{\psi_{tc}}_{\textit{baseline}} + \underbrace{\rho_c^\top \left(\frac{1}{i-1} \sum_{j=1}^{i-1} \alpha_{y_{tj}} \right)}_{\text{item interactions}}$$

- So far, the baseline captures customer preferences,

$$\psi_{tc} = \theta_u^\top \alpha_c$$

- We include extra terms,

$$\psi_{tc} = \underbrace{\lambda_c}_{\text{intercept}} + \underbrace{\theta_u^\top \alpha_c}_{\substack{\text{user} \\ \text{preferences}}} - \underbrace{\gamma_u^\top \beta_c}_{\substack{\text{price} \\ \text{sensitivity}}} \cdot \underbrace{\log(r_{tc})}_{\substack{\text{normalized} \\ \text{log-price}}} + \underbrace{\delta_w^\top \mu_c}_{\substack{\text{seasonal} \\ \text{effects}}}$$

Price and Seasonal Effects are Additive Components

$$\psi_{tc} = \underbrace{\lambda_c}_{\text{intercept}} + \underbrace{\theta_u^\top \alpha_c}_{\substack{\text{user}\\ \text{preferences}}} - \underbrace{\gamma_u^\top \beta_c}_{\substack{\text{price}\\ \text{sensitivity}}} \cdot \underbrace{\log(r_{tc})}_{\substack{\text{normalized}\\ \text{log-price}}} + \underbrace{\delta_w^\top \mu_c}_{\substack{\text{seasonal}\\ \text{effects}}}$$

- Price sensitivities are factorized (user/item factorization)
  - Normalized price
  - We constrain γ_u and β_c to be positive $\implies$ Negative elasticities
- Seasonal effects are factorized (week/item factorization)

Thinking One-Step Ahead

- Customers consider step $i+1$ when making the decision about step i
- This allows capturing complementarity (details on next slide)
- Mathematically,

$$\Psi(c, \mathbf{y}_{t,i-1}) = \psi_{tc} + \rho_c^\top \left(\frac{1}{i-1} \sum_{j=1}^{i-1} \alpha_{y_{tj}} \right) + \max_{c' \notin [\mathbf{y}_{t,i-1}, c]} \left\{ \psi_{tc'} + \rho_{c'}^\top \left(\frac{1}{i} \left(\alpha_c + \sum_{j=1}^{i-1} \alpha_{y_{tj}} \right) \right) \right\}$$

Illustrative Simulation

Consider the following world:

- There are 8 items:

coffee *diapers* *ramen* *candy* *hot dogs* *buns* *taco shells* *seasoning*

- Two types of customers:
  - New parents frequently buy *coffee* and *diapers*
  - College students frequently buy *ramen* and *candy*

- Each customer also buys either (*hot dogs*, *hot dog buns*) or (*taco shells*, *taco seasoning*)

- Customers are sensitive to price
  - Decisions about preferred items are independent
  - Decisions about complementary items are by pairs

Illustrative Simulation

purchased items: *diapers*, *hot dogs*, *hot dog buns*, *checkout*

		stage 1: *diapers*
non think-ahead	diapers	**0.31**
	coffee (↑)	0.03
	ramen	0.00
	candy	0.00
	hot dogs	0.18
	hot dog buns	0.17
	taco shells (↑)	0.14
	taco seasoning	0.17
	checkout	0.00
think-ahead	diapers	**0.37**
	coffee (↑)	0.02
	ramen	0.00
	candy	0.00
	hot dogs	0.24
	hot dog buns	0.24
	taco shells (↑)	0.06
	taco seasoning	0.06
	checkout	0.00

Illustrative Simulation

purchased items: *diapers*, *hot dogs*, *hot dog buns*, *checkout*

		stage 1: *diapers*	stage 2: *hot dogs*
non think-ahead	diapers	**0.31**	0.00
	coffee (↑)	0.03	0.02
	ramen	0.00	0.00
	candy	0.00	0.00
	hot dogs	0.18	**0.25**
	hot dog buns	0.17	0.25
	taco shells (↑)	0.14	0.19
	taco seasoning	0.17	0.24
	checkout	0.00	0.05
think-ahead	diapers	**0.37**	0.00
	coffee (↑)	0.02	0.02
	ramen	0.00	0.00
	candy	0.00	0.00
	hot dogs	0.24	**0.34**
	hot dog buns	0.24	0.42
	taco shells (↑)	0.06	0.10
	taco seasoning	0.06	0.10
	checkout	0.00	0.02

Illustrative Simulation

purchased items: *diapers*, *hot dogs*, <u>*hot dog buns*</u>, *checkout*

		stage 1: *diapers*	stage 2: *hot dogs*	stage 3: *hot dog buns*
non think-ahead	diapers	**0.31**	0.00	0.00
	coffee (↑)	0.03	0.02	0.05
	ramen	0.00	0.00	0.00
	candy	0.00	0.00	0.00
	hot dogs	0.18	**0.25**	0.00
	hot dog buns	0.17	0.25	**0.79**
	taco shells (↑)	0.14	0.19	0.00
	taco seasoning	0.17	0.24	0.00
	checkout	0.00	0.05	0.16
think-ahead	diapers	**0.37**	0.00	0.00
	coffee (↑)	0.02	0.02	0.07
	ramen	0.00	0.00	0.00
	candy	0.00	0.00	0.00
	hot dogs	0.24	**0.34**	0.00
	hot dog buns	0.24	0.42	**0.85**
	taco shells (↑)	0.06	0.10	0.00
	taco seasoning	0.06	0.10	0.00
	checkout	0.00	0.02	0.08

Illustrative Simulation

purchased items: *diapers*, *hot dogs*, *hot dog buns*, <u>*checkout*</u>

		stage 1: *diapers*	stage 2: *hot dogs*	stage 3: *hot dog buns*	stage 4: *checkout*
non think-ahead	diapers	**0.31**	0.00	0.00	0.00
	coffee (↑)	0.03	0.02	0.05	0.21
	ramen	0.00	0.00	0.00	0.00
	candy	0.00	0.00	0.00	0.00
	hot dogs	0.18	**0.25**	0.00	0.00
	hot dog buns	0.17	0.25	**0.79**	0.00
	taco shells (↑)	0.14	0.19	0.00	0.00
	taco seasoning	0.17	0.24	0.00	0.00
	checkout	0.00	0.05	0.16	**0.79**
think-ahead	diapers	**0.37**	0.00	0.00	0.00
	coffee (↑)	0.02	0.02	0.07	0.10
	ramen	0.00	0.00	0.00	0.00
	candy	0.00	0.00	0.00	0.00
	hot dogs	0.24	**0.34**	0.00	0.00
	hot dog buns	0.24	0.42	**0.85**	0.00
	taco shells (↑)	0.06	0.10	0.00	0.00
	taco seasoning	0.06	0.10	0.00	0.00
	checkout	0.00	0.02	0.08	**0.90**

Model Estimation: Bayesian Inference

- Prior on latent variables $p(\ell)$ (Gaussian+Gamma)
- Latent variables ℓ: user preferences θ_u, item attributes α_c, item intercepts λ_c, item interaction coefficients ρ_c, seasonal effect parameters δ_w and μ_c, price sensitivity parameters γ_u and β_c
- Posterior of latent variables given data,

$$p(\boldsymbol{\ell} \,|\, \mathcal{Y}, \mathbf{x}) = \frac{p(\boldsymbol{\ell}) \prod_{t=1}^{T} p(\mathcal{Y}_t \,|\, \boldsymbol{\ell}, x_t)}{p(\mathcal{Y} \,|\, \mathbf{x})}$$

Variational Inference Approximates the Posterior

- Approximate the posterior $p(\boldsymbol{\ell} \mid \mathcal{Y}, \mathbf{x})$
- Variational inference
- Introduce an approximating distribution $q(\ell)$ over the latent variables
- Find $q(\ell)$ by minimizing the KL divergence to the exact posterior

Variational Inference as an Optimization Problem

- Parameterized family of distributions $q_\nu(\ell)$
- Minimizing the KL $\equiv$ Maximizing the ELBO

$$\mathcal{L}(\nu) = \mathbb{E}_{q(\boldsymbol{\ell};\nu)}\left[\log p(\boldsymbol{\ell}, \mathcal{Y} \mid \mathbf{x}) - \log q(\boldsymbol{\ell};\nu)\right]$$

- Solve the optimization problem w.r.t. ν

$$\nu^\star = \arg\max_\nu \mathcal{L}(\nu)$$

A Sketch on the Variational Inference Algorithm

- Gradient-based stochastic optimization w.r.t. ν
  - Large datasets
  - Intractable expectations

- Variational bounds on the ELBO
  - Unordered baskets
  - Large number of items

The Dataset in Numbers

- 97 (unique) weeks of shopping data from a large grocery store
  - 570K baskets
  - 6M purchases
  - 5.5K unique items
  - 3K customers

- Split into training, test, validation
  - Training: Weeks 1-88
  - Test: Weeks 89-97
  - Validation: 5% of training

Models we Compare

Comparisons:

- Exponential family embeddings
- Hierarchical Poisson factorization
- (Multinomial logistic regression / Factor analysis)

Model	Data	User preferences	Item-to-item interactions	Price	Seasonal effects
B-Emb (Rudolph et al., 2016)	Binary	×	✓	×	×
P-Emb (Rudolph et al., 2016)	Count	×	✓	×	×
HPF (Gopalan, Hofman and Blei, 2015)	Count	✓	×	×	×
SHOPPER (this paper)	Binary	✓	✓	✓	✓

Predictions on the Test Set

Predictive log-likelihood for category-level data:

Model	Log-likelihood All (540K)	Price±2.5% (231K)	Price±5% (139K)	Price±15% (25K)
B-Emb (Rudolph et al., 2016)	−5.119 (0.001)	−5.119 (0.002)	−5.148 (0.002)	−5.250 (0.006)
P-Emb (Rudolph et al., 2016)	−5.160 (0.001)	−5.138 (0.002)	−5.204 (0.002)	−5.311 (0.005)
HPF (Gopalan, Hofman and Blei, 2015)	−4.914 (0.002)	−4.931 (0.002)	−4.994 (0.003)	−5.061 (0.009)
SHOPPER (I+U)	−4.744 (0.002)	−4.743 (0.003)	−4.776 (0.003)	−4.82 (0.01)
SHOPPER (I+U+S)	−4.730 (0.002)	−4.778 (0.003)	−4.801 (0.004)	−4.83 (0.01)
SHOPPER (I+U+P)	−4.728 (0.002)	−4.753 (0.003)	**−4.747** (0.004)	−4.69 (0.01)
SHOPPER (I+U+P+S)	**−4.724** (0.002)	**−4.741** (0.003)	−4.774 (0.004)	**−4.64** (0.01)

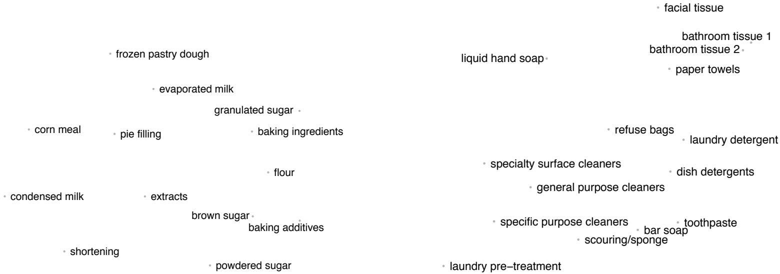
Predictions on the Test Set

Predictive log-likelihood for category-level data:

	Three items	Entire baskets
Non think-ahead	-4.795 (0.005)	-4.96 (0.02)
Think-ahead	**-4.747** (0.004)	**-4.91** (0.02)

Qualitative Results on Category-Level Data

Projected item features α_c (two regions):



Qualitative Results on Category-Level Data

Item similarities (cosine distance in α_c-space):

mollusks	organic vegetables	granulated sugar	cat food dry/moist
finfish all other - frozen	organic fruits	flour	cat food wet
crustacean non-shrimp	citrus	baking ingredients	cat litter & deodorant
shrimp family	cooking vegetables	brown sugar	pet supplies

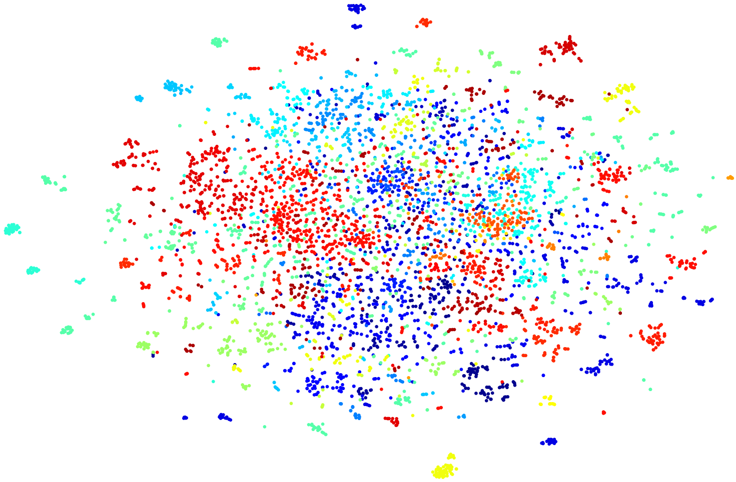
Qualitative Results on Category-Level Data

Seasonal effects (product $\delta_w^\top \mu_c$):

Halloween candy		cherries		turkey - frozen	
3.46	2006/10/25	3.07	2006/06/28	3.56	2005/11/16
3.34	2005/10/26	3.01	2006/07/12	3.30	2006/11/15
2.81	2005/10/19	2.85	2006/06/21	2.64	2005/11/23
			⋮		
−1.28	2005/11/23	−3.59	2006/10/11	−1.25	2006/06/21
−1.31	2007/01/03	−3.89	2006/10/18	−1.29	2006/07/05
−1.33	2005/11/16	−4.54	2006/10/25	−1.30	2006/07/19

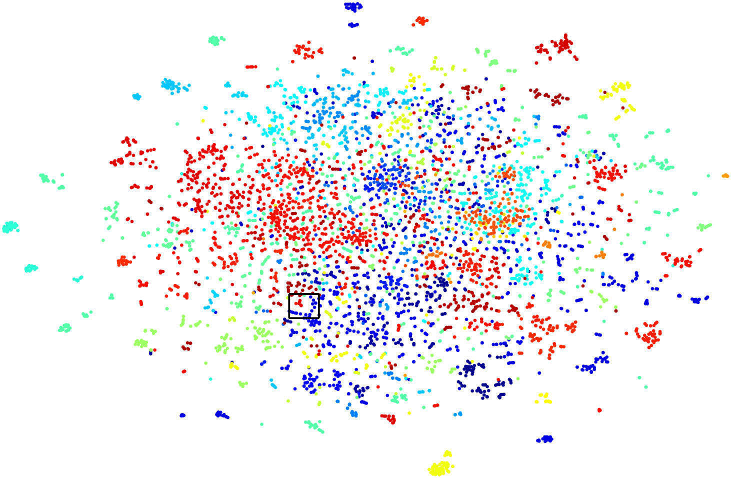
Qualitative Results on UPC-Level Data

Projected item features α_c:



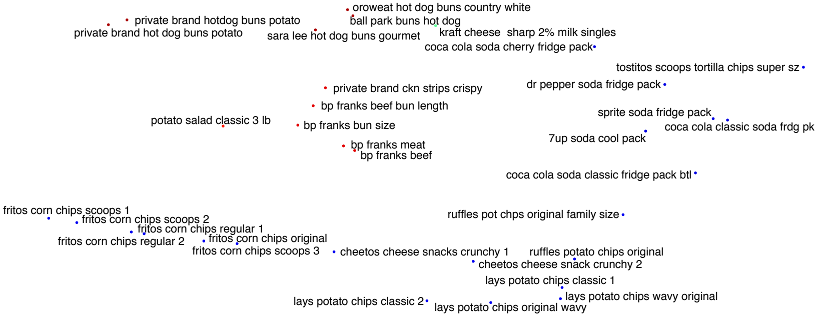
Qualitative Results on UPC-Level Data

Projected item features α_c:



Qualitative Results on UPC-Level Data

Projected item features α_c (zoom):



Complements and Substitutes

- Complementarity metric,

$$C_{cc'} \triangleq \frac{1}{2}\left(\rho_c^\top \alpha_{c'} + \rho_{c'}^\top \alpha_c\right)$$

- Exchangeability metric,

$$\begin{aligned} E_{cc'} &\triangleq \frac{1}{2}\left(D_{\mathrm{KL}}\left(p_{\cdot|c} \,||\, p_{\cdot|c'}\right) + D_{\mathrm{KL}}\left(p_{\cdot|c'} \,||\, p_{\cdot|c}\right)\right) \\ &= \frac{1}{2}\sum_{k\neq c,c'}\left(p_{k|c}\log\frac{p_{k|c}}{p_{k|c'}} + p_{k|c'}\log\frac{p_{k|c'}}{p_{k|c}}\right) \end{aligned}$$

Complements and Substitutes on UPC-Level Data

Complementarity and exchangeability metrics:

query items	complementarity score		exchangeability score	
mission tortilla soft taco 1	2.40	taco bell taco seasoning mix	0.05	mission fajita size
	2.26	mcrmck seasoning mix taco	0.07	mission tortilla soft taco 2
	2.24	lawrys taco seasoning mix	0.13	mission tortilla fluffy gordita
private brand hot dog buns	2.99	bp franks meat	0.11	ball park buns hot dog
	2.63	bp franks bun size	0.13	private brand hotdog buns potato 1
	2.37	bp franks beed bun length	0.15	private brand hotdog buns potato 2
private brand mustard squeeze bottle	0.50	private brand hot dog buns	0.15	frenchs mustard classic yellow squeeze
	0.41	private brand cutlery full size forks	0.16	frenchs mustard classic yellow squeezed
	0.24	best foods mayonnaise squeeze	0.21	heinz ketchup squeeze bottle
private brand napkins all occasion	0.78	private brand selection plates 6 7/8 in	0.09	vnty fair napkins all occasion 1
	0.50	private brand selection plates 8 3/4 in	0.11	vnty fair napkins all occasion 2
	0.49	private brand cutlery full size forks	0.12	private brand selection premium napkins

Conclusions

- SHOPPER: A probabilistic model of consumer behavior
- Posterior inference to estimate latent attributes
  - Customer preferences
  - Item attributes
  - Item-item interactions
  - Price sensitivities
  - Seasonal effects
- Interpretable model
  - Predictions under price interventions
  - Find complements and substitutes
- Code publicly available[1]

[1] `https://github.com/franrruiz/shopper-src`

Future work

- Other heuristics for utility maximization over entire baskets
- Within-basket heterogeneity
- Taste for variety
- Extensions of the thinking-ahead procedure

Thank you for your attention!