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

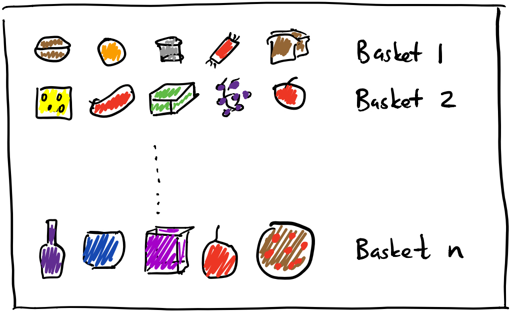
Francisco J. R. Ruiz

With Susan Athey and David M. Blei

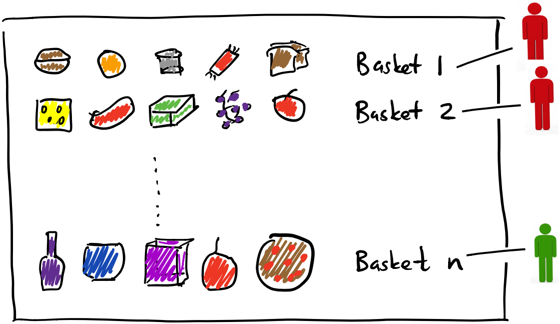
April 28th, 2020

Large-Scale Market Basket Data

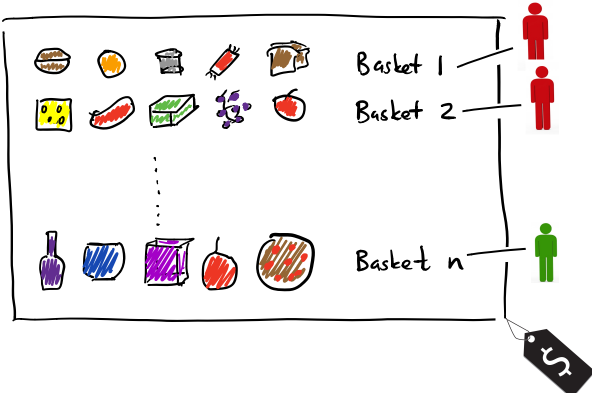
Large-Scale Market Basket Data



Large-Scale Market Basket Data



Large-Scale Market Basket Data

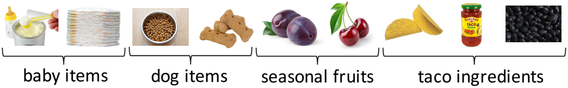


Why Analyzing Market Basket Data?

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

Market Basket Data is Complex

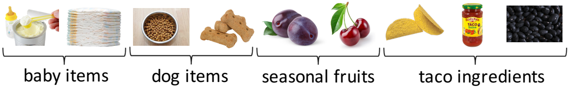
Example:



- Customer preferences vs. complements
- Seasonality effects
- Price effects

Market Basket Data is Complex

Example:



- Customer preferences vs. complements
- Seasonality effects
- Price effects

Challenges:

- Many unobserved and interrelated forces at play
- Exponential number of choices (2^C)
- Large scale: 6M baskets, 5.6K items

SHOPPER

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

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
- The sequence ends when customer chooses the "checkout item"
- The joint distribution of trip t is

$$p(\mathbf{y}_t) = p(y_{t1})p(y_{t2} \mid y_{t1}) \cdots p(y_{tn} \mid \underbrace{\mathbf{y}_t^{[n-1]}}_{\substack{\text{all previous}\\ \text{items in}\\ \text{basket}}})$$

Model: Unobserved Item/User Attributes

- The probability of choosing an item depends on *latent factors*
- Item attributes: $\alpha_c \in \mathbb{R}^K$
- Item interaction coefficients: $\rho_c \in \mathbb{R}^K$
- User preferences: $\theta_u \in \mathbb{R}^K$

SHOPPER (Vanilla Version)

$$p(\mathbf{y}_t) = p(y_{t1})p(y_{t2} \mid y_{t1}) \cdots p(y_{tn} \mid \mathbf{y}_t^{[n-1]})$$

- Probability for each step i in trip t,

$$p(y_{ti} = c \mid \underbrace{\mathbf{y}_t^{[i-1]}}_{\substack{\text{items in} \\ \text{basket}}}) = \frac{\exp\{\Psi_{tic}\}}{\sum_{c' \notin \mathbf{y}_t^{[i-1]}} \exp\{\Psi_{tic'}\}}$$

- Log-linear model with (mean) utilities

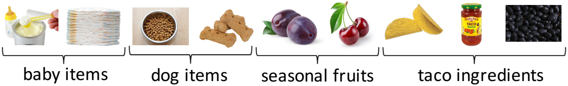
$$\Psi_{tic} = \underbrace{\psi_{tc}}_{\substack{\text{user preferences:} \\ \theta_u^\top \alpha_c}} + \underbrace{\rho_c^\top \left(\frac{1}{i-1} \sum_{j=1}^{i-1} \alpha_{y_{tj}} \right)}_{\text{item interactions}}$$

SHOPPER (Vanilla Version)

SHOPPER combines ideas from

- Matrix factorization
- Word embeddings & Exponential family embeddings

$$\Psi_{tic} = \underbrace{\psi_{tc}}_{\substack{\text{user preferences:}\\ \theta_u^\top \alpha_c}} + \underbrace{\rho_c^\top \left(\frac{1}{i-1}\sum_{j=1}^{i-1}\alpha_{y_{tj}}\right)}_{\text{item interactions}}$$

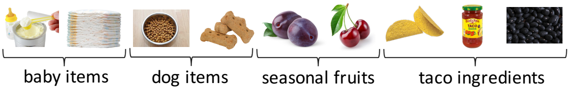


SHOPPER (Vanilla Version)

SHOPPER captures

- Customer preferences: Baby items (high $\theta_u^\top \alpha_{\text{diapers}}$)
- Complements: Taco shells and taco seasoning (high $\rho_{\text{shells}}^\top \alpha_{\text{seasoning}}$)
- Substitutes: Two brands of taco shells (low $\rho_{\text{shells}_1}^\top \alpha_{\text{shells}_2}$)

$$\Psi_{tic} = \underbrace{\psi_{tc}}_{\substack{\text{user preferences:}\\ \theta_u^\top \alpha_c}} + \underbrace{\rho_c^\top \left(\frac{1}{i-1}\sum_{j=1}^{i-1}\alpha_{y_{tj}}\right)}_{\text{item interactions}}$$



Price and Seasonal Effects

$$\Psi_{tic} = \underbrace{\psi_{tc}}_{baseline} + \underbrace{\rho_c^\top \left(\frac{1}{i-1} \sum_{j=1}^{i-1} \alpha_{y_{tj}} \right)}_{\text{item interactions}}$$

- Full SHOPPER model:

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

- Price sensitivities are factorized (user/item)
- Seasonal effects are factorized (week/item)

Price and Seasonal Effects

$$\Psi_{tic} = \underbrace{\psi_{tc}}_{baseline} + \underbrace{\rho_c^\top \left(\frac{1}{i-1} \sum_{j=1}^{i-1} \alpha_{y_{tj}} \right)}_{\text{item interactions}}$$

- Full SHOPPER model:

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

- Price sensitivities are factorized (user/item)
- Seasonal effects are factorized (week/item)

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,

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

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 $\ell = [\theta_u, \alpha_c, \lambda_c, \rho_c, \delta_w, \mu_c, \gamma_u, \beta_c]$
- Posterior of latent variables given data,

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

Variational Inference Approximates the Posterior

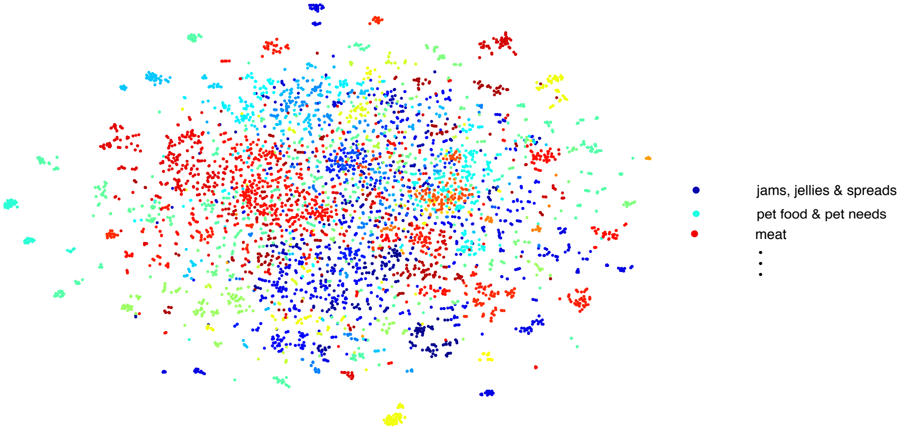
- Approximate the posterior $p(\boldsymbol{\ell} \,|\, \mathcal{Y}, \mathbf{x})$
- Variational inference
- Introduce an approximating distribution $q(\ell)$ over the latent variables
- Find $q(\ell)$ by minimizing the KL divergence to the exact posterior

The Dataset in Numbers

- 6,000,000 items
- 570,000 baskets
- 3,200 customers
- 5,600 unique items
- 2 years of data

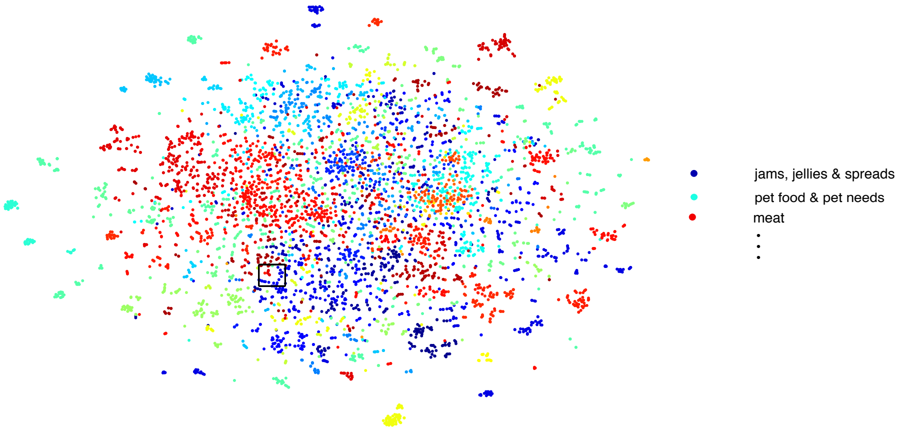
Qualitative Results

Projected item features α_c:



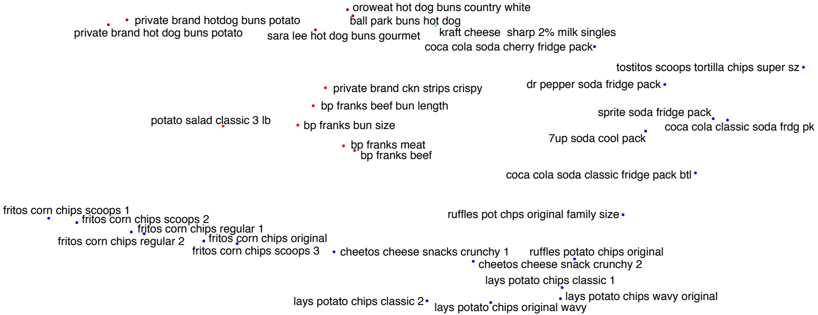
Qualitative Results

Projected item features α_c:



Qualitative Results

Projected item features α_c (zoom):



Qualitative Results

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

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
	$\vdots$		$\vdots$		$\vdots$

Predictions on the Test Set

Log-likelihood on the test set (last 8 weeks of data):

Model	Log-likelihood All (540K)	Price±2.5% (231K)	Price±5% (139K)	Price±15% (25K)
B-Emb [Rudolph+, 2016]	−5.12	−5.12	−5.15	−5.25
P-Emb [Rudolph+, 2016]	−5.16	−5.14	−5.20	−5.31
HPF [Gopalan+, 2015]	−4.91	−4.93	−4.99	−5.06
SHOPPER (I+U)	−4.74	−4.74	−4.78	−4.82
SHOPPER (I+U+S)	−4.73	−4.78	−4.80	−4.83
SHOPPER (I+U+P)	−4.73	−4.75	**−4.75**	−4.69
SHOPPER (I+U+P+S)	**−4.72**	**−4.74**	−4.77	**−4.64**

Predictions on the Test Set

Log-likelihood on the test set (last 8 weeks of data):

Model	Log-likelihood All (540K)	Price±2.5% (231K)	Price±5% (139K)	Price±15% (25K)
B-Emb [Rudolph+, 2016]	−5.12	−5.12	−5.15	−5.25
P-Emb [Rudolph+, 2016]	−5.16	−5.14	−5.20	−5.31
HPF [Gopalan+, 2015]	−4.91	−4.93	−4.99	−5.06
SHOPPER (I+U)	−4.74	−4.74	−4.78	−4.82
SHOPPER (I+U+S)	−4.73	−4.78	−4.80	−4.83
SHOPPER (I+U+P)	−4.73	−4.75	**−4.75**	−4.69
SHOPPER (I+U+P+S)	**−4.72**	**−4.74**	−4.77	**−4.64**

Good predictions on skewed test sets

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,

$$E_{cc'} \triangleq \frac{1}{2}\left(D_{\mathrm{KL}}\left(p_{\cdot|c} \;||\; p_{\cdot|c'}\right) + D_{\mathrm{KL}}\left(p_{\cdot|c'} \;||\; p_{\cdot|c}\right)\right)$$

query	complementarity score		exchangeability score	
mission tortilla taco 1	2.40	taco bell seasoning mix	0.05	mission fajita
	2.26	mcrmck seasoning mix	0.07	mission tortilla taco 2
	2.24	lawrys seasoning mix	0.13	mission tortilla fluffy gordita
(private) hot dog buns	2.99	bp franks meat	0.11	ball park hot dog buns
	2.63	bp franks bun size	0.13	(private) hot dog potato buns 1
	2.37	bp franks beed bun length	0.15	(private) hot dog potato buns 2

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,

$$E_{cc'} \triangleq \frac{1}{2}\left(D_{\mathrm{KL}}\left(p_{\cdot|c} \,||\, p_{\cdot|c'}\right) + D_{\mathrm{KL}}\left(p_{\cdot|c'} \,||\, p_{\cdot|c}\right)\right)$$

query	complementarity score		exchangeability score	
mission tortilla taco 1	2.40	taco bell seasoning mix	0.05	mission fajita
	2.26	mcrmck seasoning mix	0.07	mission tortilla taco 2
	2.24	lawrys seasoning mix	0.13	mission tortilla fluffy gordita
(private) hot dog buns	2.99	bp franks meat	0.11	ball park hot dog buns
	2.63	bp franks bun size	0.13	(private) hot dog potato buns 1
	2.37	bp franks beed bun length	0.15	(private) hot dog potato buns 2

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,

$$E_{cc'} \triangleq \frac{1}{2}\left(D_{\mathrm{KL}}\left(p_{\cdot|c} \mid\mid p_{\cdot|c'}\right) + D_{\mathrm{KL}}\left(p_{\cdot|c'} \mid\mid p_{\cdot|c}\right)\right)$$

query	complementarity score		exchangeability score	
mission tortilla taco 1	2.40	taco bell seasoning mix	0.05	mission fajita
	2.26	mcrmck seasoning mix	0.07	mission tortilla taco 2
	2.24	lawrys seasoning mix	0.13	mission tortilla fluffy gordita
(private) hot dog buns	2.99	bp franks meat	0.11	ball park hot dog buns
	2.63	bp franks bun size	0.13	(private) hot dog potato buns 1
	2.37	bp franks beed bun length	0.15	(private) hot dog potato buns 2

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,

$$E_{cc'} \triangleq \frac{1}{2}\left(D_{\mathrm{KL}}\left(p_{\cdot|c} \;||\; p_{\cdot|c'}\right) + D_{\mathrm{KL}}\left(p_{\cdot|c'} \;||\; p_{\cdot|c}\right)\right)$$

query	complementarity score		exchangeability score	
mission tortilla taco 1	2.40	taco bell seasoning mix	0.05	mission fajita
	2.26	mcrmck seasoning mix	0.07	mission tortilla taco 2
	2.24	lawrys seasoning mix	0.13	mission tortilla fluffy gordita
(private) hot dog buns	2.99	bp franks meat	0.11	ball park hot dog buns
	2.63	bp franks bun size	0.13	(private) hot dog potato buns 1
	2.37	bp franks beed bun length	0.15	(private) hot dog potato buns 2

Conclusions

- SHOPPER: A probabilistic model of consumer behavior
  - Discrete choice model with interpretable components
  - Efficient inference algorithm
  - Predictions under price interventions
  - Identify limitations

- Code publicly available[1]

EU H2020 (MSCA Actions,
Grant Agreement 706760)

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

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

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

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)

A Sketch of VI for SHOPPER

- Objective

$$\mathcal{L}(\nu) = \mathbb{E}_{q_\nu(\boldsymbol{\ell})}\left[\log p(\boldsymbol{\ell}, \mathcal{D}) - \log q_\nu(\boldsymbol{\ell})\right]$$

- Monte Carlo gradient estimator

$$\nabla_\nu \mathcal{L}(\nu) = \mathbb{E}_{q_\nu(\boldsymbol{\ell})}\left[f(\boldsymbol{\ell}, \nu)\right] \approx \frac{1}{S}\sum_{s=1}^{S} f(\boldsymbol{\ell}^{(s)}, \nu), \qquad \boldsymbol{\ell}^{(s)} \sim q_\nu(\boldsymbol{\ell})$$

- Stochastic optimization addresses
  - Large datasets
  - Intractable expectations
- Variational bounds on the ELBO
  - Unordered baskets
  - Large number of items