Exploring Complementarity: A Survey

Agenda

1. Context
2. Paper Discussion (**X** 3)
    a. Setup: Problem the Paper is Addressing
    b. Loss function/Likelihood
    c. Datasets/Training Instances/Training Methodology
    d. Inference and Results
    e. Novelty/Other
    f. Application at Etsy
3. Possible Future Avenues

Context of Survey

1. Definitions:
   a. Complementary items: buy knitting needles → buy yarn.
   b. Substitutable items: view dress A -> buy dress B.
2. Cart Recommendations at Etsy use complementary items:
   a. Complementarity information: Item to item matrix with # of times row item co-purchased with column item.
   b. Candidate selection.
   c. Feature: Cosine similarity between 2 rows in the matrix.
3. Need dedicated “People who purchased this also purchased that” (aka cross-sell) recommendation module. Current module is more general purpose.
4. Literature survey: generate ideas for new top level models for complementary items/ improve features / improve candidate selection.

Paper Setup

Inferring Networks of Substitutable and Complementary Products
[Spectre], J. McAuley et al, 2015

1. Predict if any two items have a complementary or substitutable relationship.
2. Predict the direction of the relationship. Buy bed → buy sheets, buy sheets → buy bed?
3. Discover topic models from text in product reviews that are good at predicting these links with direction.

Likelihood

Logically has two factors:

$$L(y, \mathcal{T} | \beta, \eta, \theta, \phi, z) = \overbrace{\prod_{d \in \mathcal{T}} \prod_{j=1}^{N_d} \theta_{z_{d,j}} \phi_{z_{d,j}, w_{d,j}}}^{\text{corpus likelihood}}$$

$$\underbrace{\prod_{g \in G} \left\{ \prod_{(i,j) \in \mathcal{E}_g} F^{\leftrightarrow}_{\beta_g}(\psi_\theta(i,j)) F^{\rightarrow}_{\eta_g}(\varphi_\theta(i,j))(1 - F^{\rightarrow}_{\eta_g}(\varphi_\theta(j,i))) \prod_{(i,j) \in \bar{\mathcal{E}}_g} (1 - F^{\leftrightarrow}_{\beta_g}(\psi_\theta(i,j))) \right\}}_{\text{accuracy of the predictors } \beta_g \text{ and } \eta_g \textit{ for the graph } g}. \quad (8)$$

Symbol	Description
d_i	document associated with an item (product) i
$\mathcal{T}$	document corpus
K	number of topics
θ_i	K-dimensional topic distribution for item i
ϕ_k	word distribution for topic k
$w_{d,j}$	j^{th} word of document d
$z_{d,j}$	topic of the j^{th} word document d
N_d	number of words in document d
$F(x)$	logistic (sigmoid) function, $1/(1+e^{-x})$
$\mathcal{E}_g$	observed edges in graph g
$\psi(i,j)$	pairwise (undirected) features for items i and j
$\varphi(i,j)$	pairwise (directed) features for items i and j
β	logistic weights associated with $\psi(i,j)$
η	logistic weights associated with $\varphi(i,j)$

- 1st factor: corpus likelihood product over all words in all documents
  - for each word, the product of probability of picking a topic times the probability of picking that specific word from that topic.

Likelihood (Contd.)

$$\prod_{g \in G} \left\{ \prod_{(i,j) \in \mathcal{E}_g} F_{\beta_g}^{\leftrightarrow}(\psi_\theta(i,j)) F_{\eta_g}^{\rightarrow}(\varphi_\theta(i,j))(1 - F_{\eta_g}^{\rightarrow}(\varphi_\theta(j,i))) \prod_{(i,j) \in \bar{\mathcal{E}}_g} (1 - F_{\beta_g}^{\leftrightarrow}(\psi_\theta(i,j))) \right\}.$$

- 2nd factor: edge detection between a pair of items
  - F_\beta: likelihood (sigmoid) of an edge between two items (over all edges \Epsilon_g)
  - F_\eta: likelihood (sigmoid) related to the direction of the edge (over all edges \Epsilon_g)
  - (1 - F_\beta): likelihood of non-edges (over all non-edges)
  - Similar pairs of sigmoids for each relationship, substitution and complementary. (g belongs to G in above)

Likelihood (Contd.)

$$\prod_{g \in G} \Big\{ \prod_{(i,j) \in \mathcal{E}_g} F^{\leftrightarrow}_{\beta_g}(\psi_\theta(i,j)) F^{\rightarrow}_{\eta_g}(\varphi_\theta(i,j))(1 - F^{\rightarrow}_{\eta_g}(\varphi_\theta(j,i)))$$

$$\prod_{(i,j) \in \bar{\mathcal{E}}_g} (1 - F^{\leftrightarrow}_{\beta_g}(\psi_\theta(i,j))) \Big\}.$$

If we have K topics, we learn K-length vector representations of the

items. The feature vectors used for F_\beta, the non-directional sigmoid,

are created by performing an element-wise product:

- $\psi_\theta(i,j) = (1, \theta_{i,1} \cdot \theta_{j,1}, \theta_{i,2} \cdot \theta_{j,2}, \ldots, \theta_{i,K} \cdot \theta_{j,K}).$

The feature vector used for the directional logistic function, F_\eta,

element-wise difference:

- $\varphi_\theta(i,j) = (1, \theta_{j,1} - \theta_{i,1}, \ldots, \theta_{j,K} - \theta_{i,K}),$

Datasets/Training Instances/Training Algorithm

- Amazon text reviews
- Datasets/Instances:
  - Edges like: 'Users who viewed x also viewed y' [+ve: substitute, -ve: complementary]
  - 'Users who viewed x eventually purchased y' [+ve: substitute, -ve: complementary]
  - 'Users who bought x also bought y' [+ve: complementary, -ve: substitute]
  - 'Users frequently bought x and y together' [+ve: complementary, -ve: substitute]
  - During training, substitute positive instances are negative instances for complementarity and vice versa.

Datasets/Training Instances/Training Algorithm

- Train model for each top level category.
- Training algorithm EM like algorithm:
  - First step assign topics randomly to all words in docs
  - Second step remaining params are learned using gradient ascent via a Hybrid LBFGS solver.
  - Go back to the E step: Update word -> topic assignments using likelihood of word occurring with that topic.

Inference/Results

- Evaluate model in terms of link prediction and ranking

- for each test edge (in each graph): $a \to b$,
  $F_\theta^{\leftrightarrow}(\psi(a,b),\beta) > 0$ ***and*** $F_\theta^{\rightarrow}(\varphi(a,b),\eta) > 0$

- Baselines:
  - Random is replacing sigmoids with random numbers between 0 and 1.
  - LDA topic modeling separately before learning rest of model.
  - CT Category tree: based on most common category co-counts.
  - CF Item to Item Collaborative Filtering.

Results (Contd.)

Category	Method	Accuracy Subst.	Accuracy Compl.	Error reduction vs. random Subst.	Error reduction vs. random Compl.
Men's Clothing	Random	60.27%	57.70%	0.0%	0.0%
	LDA	70.62%	65.95%	26.05%	19.50%
	CT	78.69%	61.06%	46.38%	7.946%
	Sceptre	96.69%	94.06%	91.67%	85.97%
Women's Clothing	Random	60.35%	56.67%	0.0%	0.0%
	LDA	70.70%	64.80%	26.11%	18.75%
	CT	81.05%	69.08%	52.21%	28.63%
	Sceptre	95.87%	94.14%	89.59%	86.47%
Music	Random	-	50.18%	-	0.0%
	LDA	-	52.39%	-	4.428%
	CT	-	57.02%	-	13.71%
	Sceptre	-	90.43%	-	80.78%
Movies	Random	-	51.22%	-	0.0%
	LDA	-	54.26%	-	6.235%
	CT	-	66.34%	-	30.99%
	Sceptre	-	85.57%	-	70.42%
Electronics	Random	69.98%	55.67%	0.0%	0.0%
	LDA	89.90%	61.90%	66.35%	14.06%
	CT	87.26%	60.18%	57.57%	10.17%
	Sceptre	95.70%	88.80%	85.69%	74.74%
Books	Random	69.93%	55.35%	0.0%	0.0%
	LDA	89.91%	60.59%	66.47%	11.75%
	CT	87.80%	66.28%	59.42%	24.49%
	Sceptre	93.76%	89.86%	79.25%	77.29%
Baby Clothes	random	62.93%	52.47%	0.0%	0.0%
	LDA	75.86%	54.73%	34.89%	4.75%
	CT	79.31%	64.56%	44.18%	25.43%
	Sceptre	92.18%	93.65%	78.91%	86.65%
Average	*Sceptre*	94.83%	90.23%	85.02%	80.33%

Table 3: Link prediction accuracy for substitute and complement links (the former are not available for the majority of Music/Movies products in our dataset). Absolute performance is shown at left, reduction in error vs. random classification at right.

Results (Contd.)

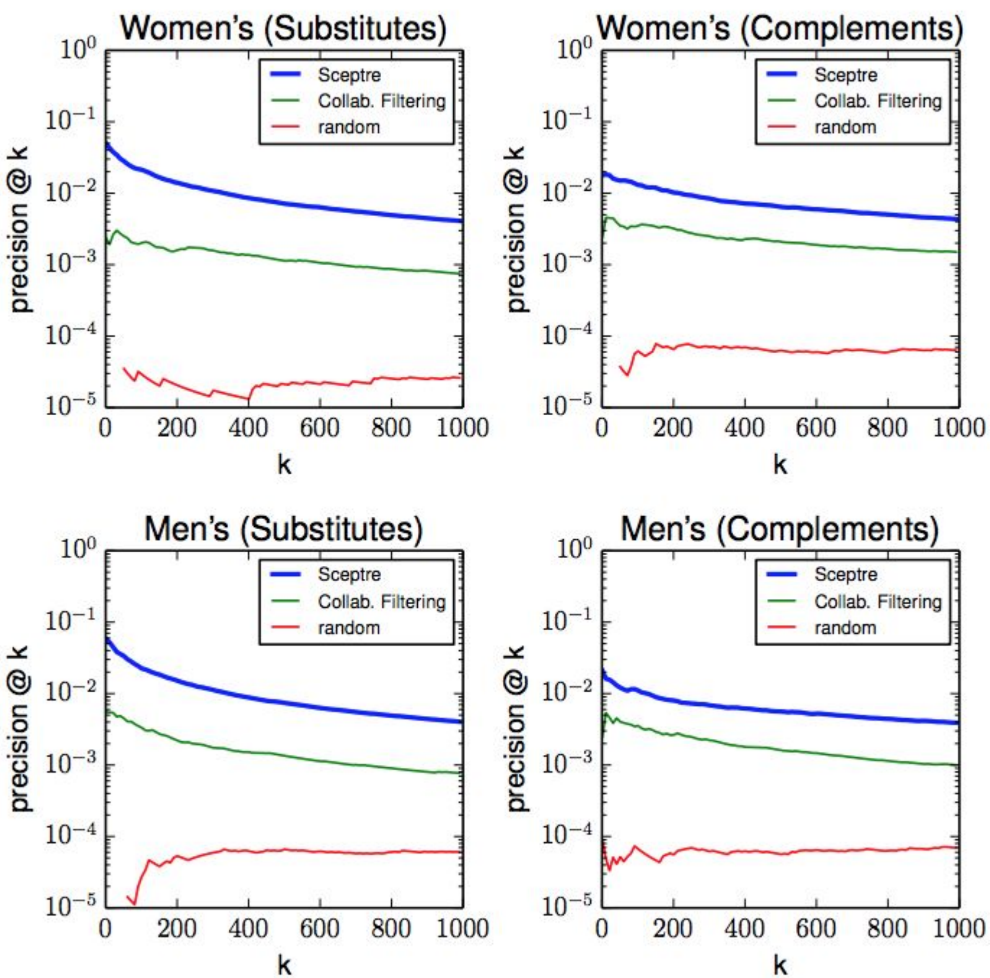


Figure 4: Precision@k for Women's and Men's clothing.

Novelty/Other

- Taxonomy tree to decide # of topics. # of topics per node = is number of unique products divided by 10,000.
- Any model for complementary items should perhaps also account for substitutable items?

Application at Etsy

- Complementarity: Ensure that we have coverage for all types of ‘edges’ described in paper, for training. (‘Users that purchased *x* also purchased *y*’, …)
- Complementarity: Pattern of modeling relationship and the direction of relationship, with sigmoids seems quite feasible/promising.
- Complementarity: Product level: module that is exclusively for recommending co-purchases/cross-sell
- General: harnessing the taxonomy for several purposes like for topic modeling, or for making graph searches more tractable.

Paper Setup

A Path-constrained Framework for Discriminating Substitutable and Complementary Products in E-commerce [Path], Z. Wang et al., 2018

- Similar to last paper, predicting substitutable and complementary linkages.
- Both whether there is a linkage between two items, and the direction of that linkage.

Loss Function

$$\min L_{joint} = L(Y|V,V') + \sum_{c=0}^{2} \alpha_c \cdot L(F_c|V,V',\boldsymbol{\beta})$$
$$= L(Y|V,V') + \alpha_0 \cdot L(Z|V,V',\boldsymbol{\beta}) + \alpha_1 \cdot L(F_1|V,V',\boldsymbol{\beta})$$
$$+ \alpha_2 \cdot L(F_2|V,V',\boldsymbol{\beta}),$$
$$\text{s.t.} \quad \|\mathbf{v_i}\|_2 \le 1 \;\; and \quad \|\mathbf{v}'_{\mathbf{i}}\|_2 \le 1, \forall i \in \mathcal{E}; \;\; \|\mathbf{r}\|_2 \le 1, \forall r \in \mathcal{R}.$$

Table 1: Glossary

Symbols	Descriptions
v_i, v'_i	target,context vector of product i
$v_{r,i}, v'_{r,i}$	projected target ,context vector of product i for r
r	product relation
β_r	projecting vector for relation r
$\mathcal{R}$	product relation set
$\mathcal{E}$	product set
$\mathcal{E}_P, \overline{\mathcal{E}}_P$	positive, negative ordered pair set
$\mathcal{E}_E, \overline{\mathcal{E}}_E$	positive, negative edge set
$\mathcal{E}_F, \overline{\mathcal{E}}_F$	positive, negative formula set
V, V'	target, context representation set
$\boldsymbol{\beta}$	projecting vector set
f, f'	positive, negative formula
N	number of negative samples

- Use dual vectors to represent each item: target vector/context vector. (v, v’)
- Represent both relationships (substitute/complementary) as a vector each; the item vectors are projected into the relationship space. (represented by \beta)

Loss Function (Contd.)

$$\begin{aligned}\min L_{joint} &= L(Y|V,V') + \sum_{c=0}^{2} \alpha_c \cdot L(F_c|V,V',\boldsymbol{\beta}) \\ &= L(Y|V,V') + \alpha_0 \cdot L(Z|V,V',\boldsymbol{\beta}) + \alpha_1 \cdot L(F_1|V,V',\boldsymbol{\beta}) \\ &+ \alpha_2 \cdot L(F_2|V,V',\boldsymbol{\beta}),\end{aligned}$$

$$\text{s.t.} \quad \|\mathbf{v_i}\|_2 \le 1 \ \ and \ \ \|\mathbf{v'_i}\|_2 \le 1, \forall i \in \mathcal{E}; \ \ \|\mathbf{r}\|_2 \le 1, \forall r \in \mathcal{R}.$$

- **4 components**. Each component has **two subcomponents**:
  - **1st sub**: sum of (negative log) sigmoids corresponding to presence of relationships
    - Sigmoid operates on the dot product of the target vector of one item with the context vector of the other item.
  - **2nd sub**: sum corresponding to the absence of relationships.
- **1st component**: ‘there is a relationship between two items’.
- **2nd component**: relationship of specific type (complementary/substitutable) between two items.
  - Sigmoid operates on the dot product of vectors projected into relationship space. Authors call this sigmoid the **‘confidence’** of a relationship.

Loss Function (Contd.)

- The confidence is written as:
  - $$I(i, r, j) = P(z_{i,j,r}) = \sigma(v_{r,i}^{T} \cdot v'_{r,j})$$
- **3rd & 4th components** capture higher level information from the data.
- **3rd component**: captures product category constraints:
  - $$\begin{aligned} &(Prod_A, RelatedTo, Prod_B) \\ \Rightarrow\quad &(Category_A, RelatedTo, Category_B). \end{aligned}$$
- Reformulated as (prod B and C are in same category):
  - $$(Prod_A, Subst, Prod_B) \Rightarrow (Prod_A, Subst, Prod_C),$$
    $$(Prod_A, Compl, Prod_B) \Rightarrow (Prod_A, Compl, Prod_C).$$
- Using “t-norm fuzzy logic”, this can be expressed in terms of the confidence terms defined above.

Loss Function (Contd.)

- **4th component** captures multi-step path constraints:
    -

$$
\begin{aligned}
(Prod_A, Subst, Prod_B) \wedge (Prod_B, Subst, Prod_C) \\
\Rightarrow (Prod_A, Subst, Prod_C), \\
(Prod_A, Compl, Prod_B) \wedge (Prod_B, Subst, Prod_C) \\
\Rightarrow (Prod_A, Compl, Prod_C), \\
(Prod_A, Subst, Prod_B) \wedge (Prod_B, Compl, Prod_C) \\
\Rightarrow (Prod_A, Compl, Prod_C), \\
(Prod_A, Compl, Prod_B) \wedge (Prod_B, Compl, Prod_C) \\
\Rightarrow (Prod_A, Compl, Prod_C).
\end{aligned}
$$

- These can also be expressed in terms of combinations of confidence terms.

Datasets/Training Instances/Training Algorithm

- Amazon and JD
- Positive instances: using same edge heuristics as previous paper.
- Negative instances generated by randomly substituting items or relationships in the positive instances.
- Train for each top-level category separately.
- Tune alpha params, learning rate, hyperparams on validation set.
- Optimization using SGD with some bells and whistles.

Inference/Results

- Evaluate link prediction and ranking
- Link Prediction:
  - if there is a directed edge of type r from i to j:
    $\sigma(v_{r,i}^T \cdot v'_{r,j}) > \alpha,$
  - \alpha is determined on a validation set

- Baselines:
  - **Random**: the confidence replaced by random numbers between 0,1. Denotes a link if greater than \alpha (selected on validation set)
  - **MF, NMF**: Matrix factorization and Non-negative MF. Elements in matrix: Edges from i->j is 1, 0 for non-links. Separate matrix for each relation.
  - **CF**: item to item collaborative filtering.
  - **Spectre**: the first paper

Results (Contd.)

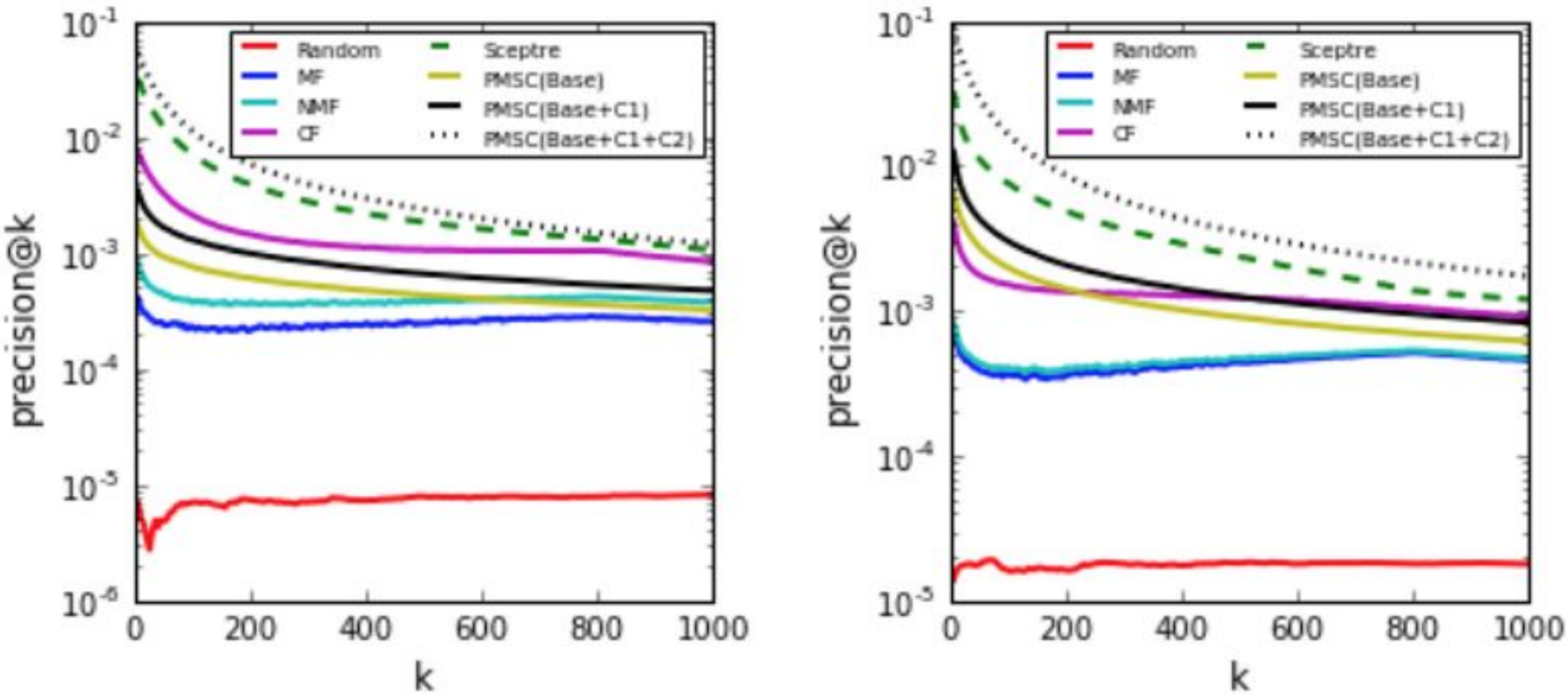
Table 5: Link Prediction Accuracy on Amazon

Method	Electronics		Women's Clothes		Home and Kitchen		Cell Phones and Accessories		Office Products	
	Subst.	Compl.	Subst.	Compl.	Subst.	Compl.	Subst.	Compl.	Subst.	Compl.
Random	0.7800	0.7801	0.7803	0.7800	0.7808	0.7796	0.7809	0.7799	0.7800	0.7804
MF	0.8056	0.7916	0.7963	0.7850	0.8140	0.8094	0.7910	0.8000	0.7950	0.7927
NMF	0.8147	0.8013	0.8179	0.8023	0.8276	0.8159	0.7979	0.8269	0.8073	0.8165
Sceptre	0.9281	0.8789	0.9142	0.9091	0.8528	0.9161	0.8966	0.8895	0.9385	0.8912
PMSC (Base)	0.8698	0.8554	0.9178	0.8595	0.8483	0.8841	0.8820	0.9068	0.8688	0.8718
PMSC (Base+C1)	0.9176	0.8735	0.9284	0.9032	0.8501	0.8973	0.8923	0.9248	0.8770	0.8508
PMSC (Base+C1+C2)	**0.9790**	**0.9242**	**0.9777**	**0.9602**	**0.8631**	**0.9252**	**0.9016**	**0.9306**	**0.9778**	**0.9079**

Table 6: Link Prediction Accuracy on JD Dataset

Method	Baby		Women's Clothes		Men's clothes		Cell Phones and Accessories		Office Products	
	Subst.	Compl.	Subst.	Compl.	Subst.	Compl.	Subst.	Compl.	Subst.	Compl.
Random	0.7799	0.7802	0.7797	0.7822	0.7797	0.7790	0.7800	0.7812	0.7799	0.7798
MF	0.7993	0.8092	0.7957	0.7961	0.7960	0.7903	0.7952	0.8072	0.8013	0.8106
NMF	0.8353	0.8070	0.8164	0.8170	0.8049	0.8281	0.8078	0.8239	0.8211	0.8284
Sceptre	0.9158	0.9156	0.8974	0.9138	0.8822	0.9105	0.9135	0.9148	0.9133	0.9154
PMSC (Base)	0.8035	0.9020	0.8761	0.7778	0.8722	0.8523	0.7691	0.8987	0.7086	0.8439
PMSC (Base+C1)	0.8174	0.9075	0.8876	0.8651	0.9104	0.8588	0.7727	0.9067	0.7283	0.9180
PMSC (Base+C1+C2)	**0.9741**	**0.9457**	**0.9411**	**0.9487**	**0.9481**	**0.9488**	**0.9570**	**0.9483**	**0.9501**	**0.9405**

Results (Contd.)



(a) Substitute (b) Complement

Figure 6: Precision@k for Cell Phones and Accessories on Amazon

Novelty/Other

- Category level and multi-step path constraints to extract more information from the same raw data. Extract previously hidden information.
- Cold start: Able to use sub-category average context and target vectors as vectors for new listings.

Application at Etsy

- Complementarity: Using the category level and multi-step path constraints seems promising.
- Complementarity: Cold Start: using average sub-category level context and target vectors for new listings, an avenue worth exploring.
- General: Having target and context level latent vector representations is worth exploring at Etsy.

Paper Setup

SHOPPER: A probabilistic model of consumer choice with substitutes and complements. F. Ruiz et al, 2017

1. Sequential probabilistic model.
2. Model the likelihood of adding an item to a shopping basket given that the basket already contains certain items.
3. The complementary and substitutable relationships are not explicitly parameterized in the model but emerge more naturally. Those relationships are not the sole focus of the paper.
4. Again, items have dual representations. Items have both **attribute vectors** and **interaction effect vectors**.

Likelihood

The probability of adding item **c** on a trip **t** as the **i*th*** item to the basket already containing **(i -1)** items can be written as a softmax probability:

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

Log unnormalized likelihood in above is modeled as:

$$\Psi(c, \mathbf{y}_{t,i-1}) = \psi_{tc} + \rho_c^\top \left(\frac{1}{i-1} \sum_{j=1}^{i-1} \alpha_{y_{tj}} \right)$$

- The second term introduces item attributes \alpha, and interaction coefficient vectors \rho for items.
- Complements here would result in a positive value for dot product between rho and alpha, substitutes would have a negative value.

Likelihood (Contd.)

- The first term captures the generative aspect of the model:

$$\psi_{tc} = \underbrace{\lambda_c}_{\text{item popularity}} + \underbrace{\theta_{u_t}^\top \alpha_c}_{\text{item interaction}} - \underbrace{\gamma_{u_t}^\top \beta_c \log r_{tc}}_{\text{price effects}} + \underbrace{\delta_{w_t}^\top \mu_c}_{\text{seasonal effects}}.$$

Datasets/Training Instances/Training Algorithm

- Multiple trips dataset with basket contents information. Unordered baskets.
- Baselines:
  - HPF hierarchical Poisson Factorization focuses on user preferences. User preferences vector, item attributes vector.
  - B-Emp: exponential family embeddings model, specifically bernoulli embeddings for binary data.
  - P-Emb: exponential family embeddings model, specifically Poisson embeddings for purchase counts. Target and context vectors.

Inference/Results

- They place priors on the latent parameters, and estimate them by calculating the approximate posterior using approximate variational inference.
- Calculate approximate posterior for each top-level category.
- In terms of results, since unordered baskets, they consider each item as the last added item to the basket, and calculate the log likelihood of adding that item to the basket. They present the average value of this log likelihood over all items in the basket, i.e. by treating each item as the last added item.

Inference/Results

Model	Log-likelihood All (320K)	Price±10% (66K)	Price±20% (20K)	Price±30% (1.5K)
B-Emb (Rudolph et al., 2016)	−5.13	−5.30	−5.33	−5.35
P-Emb (Rudolph et al., 2016)	−5.13	−5.34	−5.42	−5.48
HPF (Gopalan, Hofman and Blei, 2015)	−4.97	−5.24	−5.35	−5.45
This paper (I+U)	−4.94	−5.21	−5.27	−5.33
This paper (I+U+P)	−4.93	−5.13	−5.09	−5.01
This paper (I+U+P+S)	**−4.92**	**−5.12**	**−5.08**	**−5.00**

TABLE 3

Average predictive log-likelihood on the test set, conditioning on the remaining items of each basket. SHOPPER *with user preferences improves over the existing models. The improvement grows when adjusting for price and seasonal effects, and especially so when using skewed test sets that emulate price intervention.*

Results (Contd.)

- Also estimate the latent parameters using MAP over all baskets independent of category.
- Define a complementarity metric in this context.
    - $$C_{cc'} \triangleq \frac{1}{2}\left(\rho_c^\top \alpha_{c'} + \rho_{c'}^\top \alpha_c\right)$$

query items	complementarity score	
mission tortilla soft taco	2.51	ortega taco shells white corn
	2.40	mcrmck seasoning mix taco
	2.26	lawrys taco seasoning mix
private brand hot dog buns	3.02	bp franks bun size
	2.94	bp franks beef bun length
	2.86	private brand hamburger buns
private brand mustard squeeze bottle	0.53	private brand hamburger buns
	0.44	private brand cutlery full size asst
	0.29	private brand hot dog buns
private brand napkins all occasion	1.01	private brand cutlery full size forks
	0.62	dixie heavy duty plates dspbl 10 1/4 in
	0.39	private brand plate dsgnr 6 7/8 in

Novelty/Other

- Lots of novelty related to the variational inference set of techniques they use.
- Like how similarity and complementary behavior emerges naturally from the model, and makes intuitive sense in the context of adding items to the basket.

Application at Etsy

- General: Dual vector representation of items like the last paper. Similar to target and context vectors.
- General: consider exploring more generative models to represent the cart at etsy.

Future Avenues

- Complementarity: Ensure we have coverage for all types of ‘edges’ described, for training. (‘Users that purchased *x* also purchased *y*’, …)
- Complementarity: Using the category level and multi-step path constraints seems promising.
- General: Having target and context level latent vector representations is worth exploring at Etsy.
- Complementarity: Pattern of modeling relationship and the direction of relationship, with sigmoids seems quite feasible/promising.
- Complementarity: Product level: module that is exclusively for recommending co-purchases/cross-sell