Partially-Ranked Choice Models for Data-Driven Assortment Optimization

Sanjay Dominik Jena
Andrea Lodi
Hugo Palmer

Canada Excellence Research Chair, andrea.lodi@polymtl.ca

CERMICS 2018 – June 29, 2018, Fréjus

Assortment planning: Context

- Process of identifying the set of products that should be offered to the customer
- Direct impact on profit
  - online ads: number of clicks on ads; sales by visiting links, etc.
  - retail: conversion rate of a product, i.e., frequency of sales

Examples:

- Online advertising

- Brick-and-mortar retail

Assortment planning: Objectives

- Find assortment that maximizes revenue
- Encourage the user to select the product(s) that has/have highest utility (e.g. profit)
- In retail: assortment changes can be quite costly

Examples:

- Online advertising

- Brick-and-mortar retail

Assortment Planning: Challenges

- Small assortments $\Longrightarrow$ less choice $\Longrightarrow$ less sales !
- More products $\Longrightarrow$ more choice $\Longrightarrow$ more sales ?
  - offering all products is known to be non-optimal
- Substitution effect
  - the presence of a product may jeopardize the sales of another
  - e.g. the Apple iPad reduced the sales of the Apple Powerbook
  - the absence of a preferred product may encourage the customer to "substitute" to a (more profitable) alternative
- Complexity of assortment constraints:
  - capacity: limited shelf size or space on website
  - product dependencies: subset constraints, balance between product categories (e.g. male and female shoe models) etc.

Assortment Planning: Challenges

Given historical data on assortments and transactions:

How to learn from historical transaction data to predict the performance of a future assortment?

→ customer choice models

Parametric Choice Models

Multinomial Logit (MNL) models

- Attributes an utility to each product
- The probability that a customer selects product i from assortment S is: $\mathbb{P}(i|S) = (e^{u_i})/(e^{u_0} + \sum_{j \in S} e^{u_j})$
- Independence of Irrelevant Alternatives (IIA) property
  - Cannot capture substitution effect

Nested Logit (NL) models

- capture certain substitution among categories, but each nest is subject to the IIA property

Mixed Multinomial Logit (MMNL) models

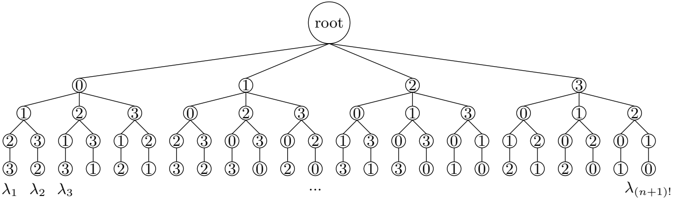
- Overcomes shortfalls of MNL and NL models
- Computationally expensive; overfitting issues

Rank-based choice models

Customer behavior σ_k: list of products ranked according to preferences of customer k, e.g. (2, 4, 0, 1, 3, 5, 6):

Customer selects highest ranked product in the assortment.

Choice model: composed of **behaviors** $\boldsymbol{\sigma}$ and corresponding **probabilities** λ_k that a random customer follows behavior σ_k.



Recent approaches using rank-based choice models

Challenge: an N-factorial large search space of customer behaviors

- Honhon et al. (2012), Vulcano and Van Ryzin (2017), ect.
  - require market knowledge, e.g. customer behaviors
- Jagabathula (2011) and Farias et al. (2013)
  - find the worst-case choice model for a given assortment
  - tractable approach to estimate probabilities for all behaviors
  - find the *sparsest* model
- Bertsimas and Misic (2016)
  - master problem minimizes estimation error for given behaviors
  - column generation to find new customer behaviors
  - pricing problem solved heuristically, since exact MIP intractable
  - limited to small number of products

Scope and Objectives of this work

Objectives:

- Develop an (efficient) data-driven approach to design optimized assortments
- Consider substitution effect (cannibalization)
- Integrate complex side constraints on the assortment (size, precedence, etc.)
- Be easy to interpret and provide market insights to management: *sparse* and *concise* models

Scope and Objectives of this work

Objectives:

- Develop an (efficient) data-driven approach to design optimized assortments
- Consider substitution effect (cannibalization)
- Integrate complex side constraints on the assortment (size, precedence, etc.)
- Be easy to interpret and provide market insights to management: *sparse* and *concise* models

Industrial collaboration:

- JDA Labs (research lab of JDA Software)
- Data from a large North-American retail chain
  - clothes (shoes and shirts)
  - seasonal choice of products

jda. Plan to deliver™

Partially-Ranked Choice Models with Indifference Sets

A new choice model:

- The customer has a strict preference on certain products.
- If unavailable, the customer may buy any *similar* product, which is available, without preference.

Partially-Ranked Choice Models with Indifference Sets

A new choice model:

- The customer has a strict preference on certain products.
- If unavailable, the customer may buy any *similar* product, which is available, without preference.

Consider a customer behavior $(P(\sigma), I(\sigma), 0)$, e.g. $(3, 4, 1, \{2, 5, 6\}, 0)$

- $P(\sigma) = (3, 4, 1) \subseteq \mathcal{N}$ is a strictly ranked list of preferred products
- $I(\sigma) = \{2, 5, 6\} \subseteq \mathcal{N} \backslash P(\sigma)$ is the subset of *indifferent* products which will be chosen with uniform probability

Partially-Ranked Choice Models: Properties

(I) Equivalence of choice models:

- Transformation from fully-ranked $(\boldsymbol{\sigma}_C, \boldsymbol{\lambda}_C)$ to partially-ranked choice model $(\boldsymbol{\sigma}_P, \boldsymbol{\lambda}_P)$, and vice versa.
- Partially-ranked behaviors more compact: factorially large number of fully-ranked behaviors required to represent the same buying behavior

Partially-Ranked Choice Models: Properties

(I) Equivalence of choice models:

- Transformation from fully-ranked $(\boldsymbol{\sigma}_C, \boldsymbol{\lambda}_C)$ to partially-ranked choice model $(\boldsymbol{\sigma}_P, \boldsymbol{\lambda}_P)$, and vice versa.
- Partially-ranked behaviors more compact: factorially large number of fully-ranked behaviors required to represent the same buying behavior

(II) (Ir)relevance of low ranked products:

- low ranked products → less important & explain less sales
  - e.g. in assortment density 0.5, the probability that product at rank 10 is selected from an "average" assortment is 0.05%
- explanatory power of indifference sets in "average" assortment is similarly low
- → concise list of strictly ranked products → insights for managers

Simplified Partially-Ranked Choice Model

Given:

- equal transformation: partial to completely ranked behaviors
- irrelevance of low ranked products and the likely small impact of indifference set on explaining the sales

we consider a **simplified** variant:

$(P(\sigma), I(\sigma), 0)$, where:

- $P(\sigma) = (3, 4, 1) \subseteq \mathcal{N}$ is a strictly ranked list of preferred products
- $I(\sigma) = \mathcal{N} \backslash P(\sigma) = \{0, 2, 5, 6\}$ is the indifference set.

$\Longrightarrow$ several computational advantages without compromising theoretical coherence

Training and Testing the Choice Models

Training set

- Set of M assortments: $\{S_m\}, m = 1, \ldots, M$
- Probabilities of selling product i in assortment S_m to a random customer: $(v_{i,m})$

Training and Testing the Choice Models

Training set

- Set of M assortments: $\{S_m\}, m = 1, \ldots, M$
- Probabilities of selling product i in assortment S_m to a random customer: $(v_{i,m})$

Test set

- Sales for each product i in each of the M other assortments

Training the choice model: completeley ranked behaviors

- Given: a subset of customer behaviors and historical sales $\boldsymbol{v}$
- Find: probability distribution ($\boldsymbol{\lambda}$) that best explains the sales
- Define a choice matrix $\boldsymbol{A}$, for each behavior k and product/assortment tuple (i, m) (BM, 2016):

$$A_{i,m}^{k} = \begin{cases} 1 \; \textit{if} \; i \text{ is chosen by customer } k \text{ in assortment } S_m \\ 0 \; \textit{if} \; i \text{ is not chosen by customer } k \text{ in assortment } S_m \end{cases}$$

Training the choice model: completeley ranked behaviors

- Given: a subset of customer behaviors and historical sales $\boldsymbol{v}$
- Find: probability distribution ($\boldsymbol{\lambda}$) that best explains the sales
- Define a choice matrix $\boldsymbol{A}$, for each behavior k and product/assortment tuple (i, m) (BM, 2016):

$$A^k_{i,m} = \begin{cases} 1 \textit{ if } i \text{ is chosen by customer } k \text{ in assortment } S_m \\ 0 \textit{ if } i \text{ is not chosen by customer } k \text{ in assortment } S_m \end{cases}$$

Linear program to find $\boldsymbol{\lambda}$ that minimizes estimation error

$$\begin{aligned} \min_{\lambda,\epsilon^+,\epsilon^-} \quad & 1^T\epsilon^+ + 1^T\epsilon^- \\ \text{s.t.} \quad & A\lambda + \epsilon^+ - \epsilon^- = v \quad (\alpha) \\ & 1^T\lambda = 1 \quad (\nu) \\ & \lambda, \epsilon^+, \epsilon^- \geq 0 \end{aligned}$$

Training the choice model: partially-ranked behaviors

Required: the total probability of buying a product sums to 1:

$$\forall (k,m), \sum_i A^k_{i,m} = 1$$

Training the choice model: partially-ranked behaviors

Required: the total probability of buying a product sums to 1:

$$\forall(k, m), \sum_i A^k_{i,m} = 1$$

For partially-ranked behaviors, a term $\frac{1}{|S_m|}$ is distributed on the products in the indifference set:

$$A^k_{i,m} = \begin{cases} 1, & \text{if } i \text{ is chosen by customer } k \text{ among assortment } S_m \\ 0, & \text{if } i \text{ is not chosen by customer } k \text{ among assortment } S_m \\ \frac{1}{|S_m|}, & \text{if } i \in S_m,\ i \in I(\sigma_k) \text{ and } P(\sigma_k) \cap S_m = \emptyset \end{cases}$$

How to efficiently find important behaviors?

- BM (2016) use column generation to find columns k for the LP above
  - MIP pricing problem is intractable for large instances; local search converges slowly

How to efficiently find important behaviors?

- BM (2016) use column generation to find columns k for the LP above
  - MIP pricing problem is intractable for large instances; local search converges slowly

Questions:

- How to exploit special structure of indifference sets?
- How to exploit the fact that high ranked products have much more impact?
  - low-ranked products may eventually not be considered

How to efficiently find important behaviors?

- BM (2016) use column generation to find columns k for the LP above
  - MIP pricing problem is intractable for large instances; local search converges slowly

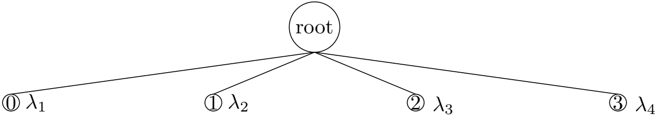
Questions:

- How to exploit special structure of indifference sets?
- How to exploit the fact that high ranked products have much more impact?
  - low-ranked products may eventually not be considered

$\Longrightarrow$ expansion of a tree: each node represents a behavior.
$\Longrightarrow$ Growing Decision Tree (GDT)

Expansion of the Growing Decision Tree (GDT)

Iteration 1



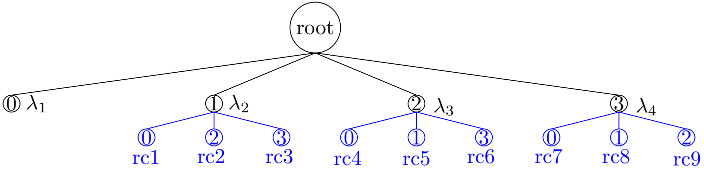
Tree is initialized with N behaviors: one for each product:

- $P(\sigma_1) = (0)$, $I(\sigma_1) = \emptyset$
- $P(\sigma_2) = (1)$, $I(\sigma_2) = \mathcal{N} \setminus 1 = \{0, 2, 3\}$
- $P(\sigma_3) = (2)$, $I(\sigma_3) = \mathcal{N} \setminus 2 = \{0, 1, 3\}$
- $P(\sigma_4) = (3)$, $I(\sigma_4) = \mathcal{N} \setminus 3 = \{0, 1, 2\}$

Master problem is solved.

Expansion of the Growing Decision Tree (GDT)

Iteration 1

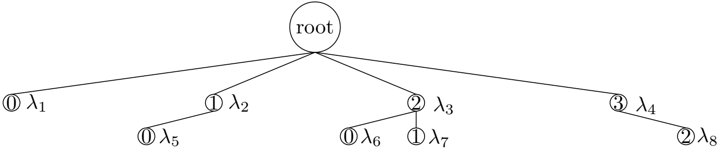


For each of the relevant customer behaviors σ_k:

- compute reduced costs (using dual values from Master problem)

Expansion of the Growing Decision Tree (GDT)

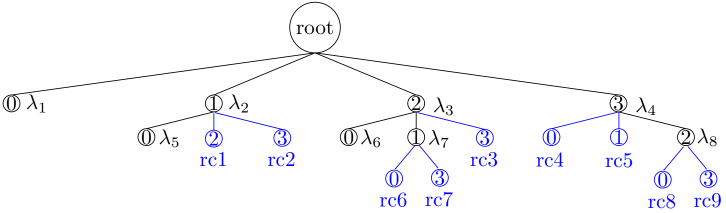
Iteration 2



- add customer behaviors with lowest (negative) reduced costs to the Master problem
- resolve Master problem

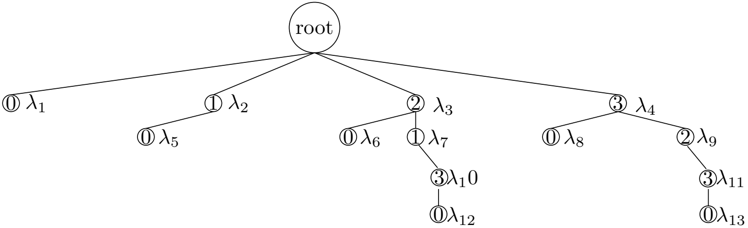
Expansion of the Growing Decision Tree (GDT)

Iteration 2



Expansion of the Growing Decision Tree (GDT)

Iteration 3



Problem Instances

Randomly generated instances:

- Via Mixed Multinomial Logit model: K classes (one for each customer type)
- Uniformly $[0, 1]$ chosen utilities for all products
- Random selection of 4 products: 100 times higher utilities
- M (typically $= 40$) assortments (20 to train, 20 to test)

Problem Instances

Randomly generated instances:

- Via Mixed Multinomial Logit model: K classes (one for each customer type)
- Uniformly $[0, 1]$ chosen utilities for all products
- Random selection of 4 products: 100 times higher utilities
- M (typically $= 40$) assortments (20 to train, 20 to test)

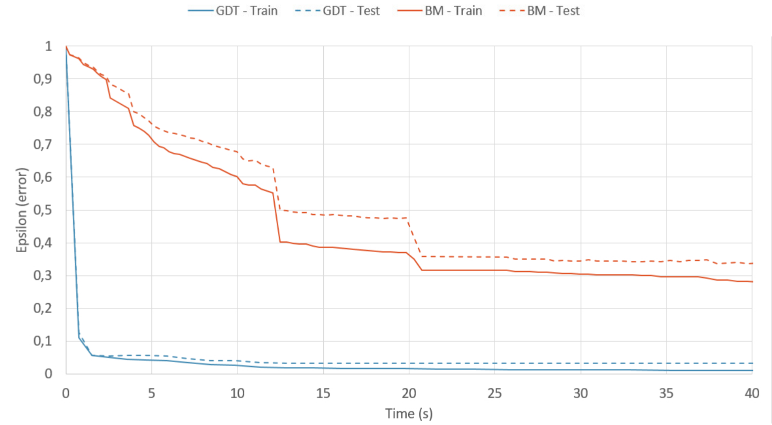
Industrial data:

- From JDA Labs: Northamerican retail chain (shoes)
- 10 stores during 10 consecutive weeks $\implies$ 100 assortments
- 192 products

Generated data - Learning curves

$n = 100$ products

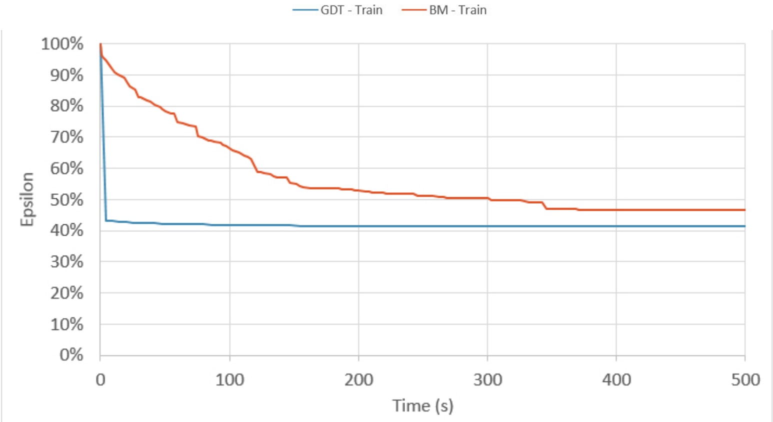
Learning curves: training and test error CG-GDT and CG-LS (BM)



Real industrial data - Computational results

$n = 192$ products

Learning curves: training and test error CG-GDT and CG-LS (BM)



Training phase: Scalability & Sparsity

Learning performance and generated choice models sizes K for:

- CG-GDT
- CG-LS

Averaged over 10 random instances
$M = 20$, $\epsilon_0 = 0.01$
Assortment density $= 0.3$ (assortment size equals $0.3 \times N$)
24 hours time limit, 48 Gbyte memory limit

	CG-GDT				CG-LS				
N	Train. error	time (sec)	# iter	K	# inst. oom	Train. error	time (sec)	# iter	K
30	0.37	2.3	9.2	105.6	0	0.39	22.5	392.0	223.8
50	0.38	6.0	10.3	104.7	0	0.40	57.3	603.2	370.1
100	0.39	29.7	15.4	127.3	0	0.40	269.8	1,070.7	721.3
250	0.39	321.8	21.0	213.3	0	0.40	5,204.8	2,492.9	1,788.7
500	0.38	2,341.5	19.4	416.6	1	0.40	49,615.3	4,555.0	3,484.2
1000	0.33	5,511.2	7.0	850.2	10	-	-	-	-
all (avg)	0.38	1,368.7	13.7	303.0	11	0.40	10,459.6	1,795.6	1,295.3

Results for assortment densities 0.1 and 0.5 show the same tendencies.

Choice model: concision

Characteristics of the generated choice model

Averaged over 10 random instances
$M = 20$, $\epsilon_0 = 0.01$
Assortment density = 0.3 (assortment size equals $0.3 \times N$)
720 minutes time limit, 48 Gbyte memory limit

ϵ_0	N	# iter	K	# strictly ranked products		% explained by indifference sets	
				avg	max	exact comp.	theor. est.
0.01	30	10.2	105.6	2.24	4	20.28	21.22
0.01	50	11.3	104.7	1.84	4	29.69	27.85
0.01	100	16.4	127.3	1.55	3	36.35	34.17
0.01	250	22.0	213.3	1.22	3	44.76	43.07
0.01	500	20.4	416.6	1.07	3	47.78	47.69
0.01	1000	8.8	836.2	1.03	2	48.54	48.98
0.01	all	14.9	300.6	1.49	4	37.90	35.59

Assortment optimization

Assortment optimization

Given a choice model, which subset of the products is likely to maximize the revenue?

Literature

- Problem **NP-hard** (2^n revenues to compute by explicit enumeration).
- *If all prices are equal*: Mahajan & van Ryzin (1999) have proposed a linear-complexity algorithm.
- General case: only heuristics (see for example *ADXOpt* by Jagabathula (2011).
- Parametric choice models generally lead to difficult formulations for assortment optimization.

Assortment Optimization: Mixed Integer Programming

Completely ranked preference lists:

- Efficient MIP to find optimal assortment (BM, 2016)
- MIP requires completely ranked customer behaviors

Assortment Optimization: Mixed Integer Programming

Completely ranked preference lists:

- Efficient MIP to find optimal assortment (BM, 2016)
- MIP requires completely ranked customer behaviors

Partially-ranked lists from GDT:

- (a) boosting: remaining ranks can be completed at random
- (b) add "indifference constraints":
  - If strictly ranked products are not in the assortment: distribute sales flow (1 unit) uniformly on all products in the indifference set that are part of the assortment
  - forces all products with equal rank to take same values
  - $K \times N^2$ constraints $\implies$ branch-and-cut

Assortment optimization: Scalability

Scalablity of assortment optimization for:

- CG-GDT with AO B&C
- CG-GDT with AO-Boosting
- CG-LS with classical AO-MIP

Averaged over 10 random instances
720 minutes time limit, 48 Gbyte memory limit

	CG-GDT with AO B&C			CG-GDT with AO-Boosting				CG-LS with classical AO-MIP			
N	K	time (min)	GT revenue	# oot oom	K	time (min)	GT revenue	# oot oom	K	time (min)	GT revenue
30	109.9	0.1	74.5	0	386.8	0.0	74.2	0	220.0	0.0	73.6
50	113.5	0.1	82.5	0	397.9	0.1	81.9	0	379.8	0.1	81.9
100	117.8	0.8	88.8	0	407.7	0.6	86.0	0	722.0	2.0	86.3
250	211.0	7.3	90.4	0	655.5	9.5	88.9	0	1,813.1	141.3	89.7
500	438.1	113.1	94.5	0	1,321.7	249.4	92.9	10	-	-	-
1000	897.4	669.9	95.0	10	-	-	-	10	-	-	-
all	314.6	131.9	87.6	10	633.9	51.9	84.8	20	783.7	35.8	82.9

Revenue: value based on ground-truth MMNL model
Boosting: at least 3 randomly completed lists for each k

Summary

- New representation for rank-based choice models
  - Indifference sets
  - Implicitly equivalent to choice models with completely ranked behaviors

Summary

- New representation for rank-based choice models
  - Indifference sets
  - Implicitly equivalent to choice models with completely ranked behaviors
- Computational advantages
  - Fast training of choice model; good convergence after few iterations
  - Fast generation of new customer behaviors (products with high ranks have more impact)

Summary

- New representation for rank-based choice models
  - Indifference sets
  - Implicitly equivalent to choice models with completely ranked behaviors
- Computational advantages
  - Fast training of choice model; good convergence after few iterations
  - Fast generation of new customer behaviors (products with high ranks have more impact)
- Advantages from the managerial perspective
  - Model is sparse: less customer behaviors
  - Model is concise: low number strictly ranked products

Open research directions

Extensions:

- Learn the choice model by "classical" ML algorithms
- Generalization to new products: how can we learn the importance of products that have never been part of past assortments?

Q (?) & A (!)

References

- Bertsimas, D., & Misic, V. (2016). Data-Driven Assortment Optimization. Working Paper. Massachusetts Institute of Technology.
- Farias, V. F., Jagabathula, S., & Shah, D. (2013). A Nonparametric Approach to Modeling Choice with Limited Data. Management Science, 59, 44.
- Honhon, D., Jonnalagedda, S., & Pan, X. A. (2012). Optimal Algorithms for Assortment Selection Under Ranking-Based Consumer Choice Models. Manufacturing & Service Operations Management, 14(2), 279-289.
- Jagabathula, S. (2011). Nonparametric Choice Modeling: Applications to Operations Management. PhD thesis. Massachusetts Institute of Technology.
- Mahajan, S. & G. J. van Ryzin. (1999). Retail inventories and customer choice. Quantitative Models for Supply Chain Management. Springer, 491-551.
- Palmer, H. (2016). Large-Scale Assortment Optimization. Master thesis. École Polytechnique de Montréal.
- Vulcano, G., & Van Ryzin, G. (2017). Technical Note - An expectation-maximization method to estimate a rank-based choice model of demand. Operations Research, 65(2), 396-407.