A Discrete Choice Model for Subset Selection

Austin R. Benson, Ravi Kumar, and Andrew Tomkins
arb@cs.cornell.edu, ravi.k53@gmail.com, atomkins@gmail.com
Code & data ⟶ https://github.com/arbenson/discrete-subset-choice

Cornell University
Research at Google

Overview: singleton vs. subset choice

Given a set of alternatives to choose from, how do people choose?

- *If choosing just one thing* (buying a car, picking a restaurant, etc.), there are lots of good ML techniques (logistic regression, your favorite deep net, etc.)
- *If choosing a subset of the alternatives* (what to buy after browsing Amazon, constructing a playlist on Spotify, etc.), there aren't as many tools.

We provide an interpretable and computationally feasible model for subset selection based on random utility maximization.

Basic concept of the model

You are throwing a small party and want to provide some snacks.
Large set of snack options and want to choose a couple.

{tortilla chips, potato chips, cookies, pretzels, guacamole, celery, nut mix, hummus, meatballs, cupcakes, pigs in blankets, cupcakes, potato skins, chicken wings, taquitos, pineapple, …}

Model 1 (Separable model).
Independent choices.
Easy computation, but not realistic.

Our model.
Some "special subsets" as options + independent choices.
Interpolate between computation and modeling power.

Model 2 (Full Model).
All subsets as options.
Harder computation, but more realistic.


today.com


healthyfoodtribe.com

Discrete choice model for subset selection

Simplest case: choices are size-2 subsets.

A person makes a selection based on random utility U_{ij} of sets *{i, j}*.

$$U_{ij} = \begin{cases} V_i + V_j + \varepsilon_{ij} & \{i,j\} \notin H \\ V_i + V_j + W_{ij} + \varepsilon_{ij} & \{i,j\} \in H, \end{cases}$$

Set of special subsets

base item utilities **corrective utility of subset**

The ε_{ij} are i.i.d. errors (per person, per choice) sampled from a Gumbel distribution.

A "rational agent" that chooses the set with largest utility chooses *{i, j}* from a set of alternatives *C* containing *i* and *j* with probability

$$\frac{p_{ij}}{\sum_{\{k,l\} \subset C} p_{kl}} \qquad p_{ij} = \begin{cases} \gamma p_i p_j & \{i,j\} \notin H \\ \gamma p_i p_j + q_{ij} & \{i,j\} \in H, \end{cases}$$

$$\sum_i p_i = 1,\ p_i \geq 0,\ \gamma \geq 0,\ p_{ij} \geq 0,\ \sum_{\{i,j\} \subset U} p_{ij} = 1$$

Generalizing to larger sets.

A person makes a selection based on random utility U_{ijk} of sets *{i, j, k}*.
Key concept. Base item utilities (the V_i) are the same regardless of size of set.

$$U_{ijk} = \begin{cases} V_i + V_j + V_k + \varepsilon_{ijk} & \{i,j,k\} \notin H \\ V_i + V_j + V_k + W_{ijk} + \varepsilon_{ijk} & \{i,j,k\} \in H, \end{cases}$$

Suppose that $H = \{\{i,j,k\}, \{i,j,k\}\}$, then
$\Pr(\text{choose } \{i,j\} \mid \text{size-2 choice}) \propto \gamma_2 p_i p_j + q_{ij}$
$\Pr(\text{choose } \{i,j,k\} \mid \text{size-3 choice}) \propto \gamma_3 p_i p_j p_k + q_{ijk}$

Putting everything together.

Use a mixture model and condition on size of selected subset.

$$\Pr(\text{select } S \mid \text{alternatives } C) = \frac{z_k}{z_1 + \cdots + z_{|C|}} \cdot \Pr(\text{select } S \mid C, \text{size-k selection}),$$

$z_k \geq 0, k = 1, \ldots, n,\quad \sum_{k=1}^{n} z_k = 1,\quad$ n = size of largest choice set

Observation. Likelihood of z_k is concave with a linear constraints ⟶ easy to learn.

Acknowledgements.
This research was partially completed while ARB was visiting Google.
ARB also supported in part by a Simons Investigator Award.

Findings with "universal choice datasets"

Universal choice datasets: the set of available alternatives is always the same.

- **Bakery.** Sets of things purchased at a bakery.
- **WalmartItems.** Sets of items bought at Walmart.
- **WalmartDepts.** Sets of departments from which items were purchased at Walmart.
- **Kosark.** Sets of hyperlinks visited during a session on a Hungarian news portal.
- **Instacart.** Sets of items in In
- **LastfmGenres.** Sets of genres of music listened to in a listening session on Last.fm.

The z_k are the faction of choices that are size-*k* sets.

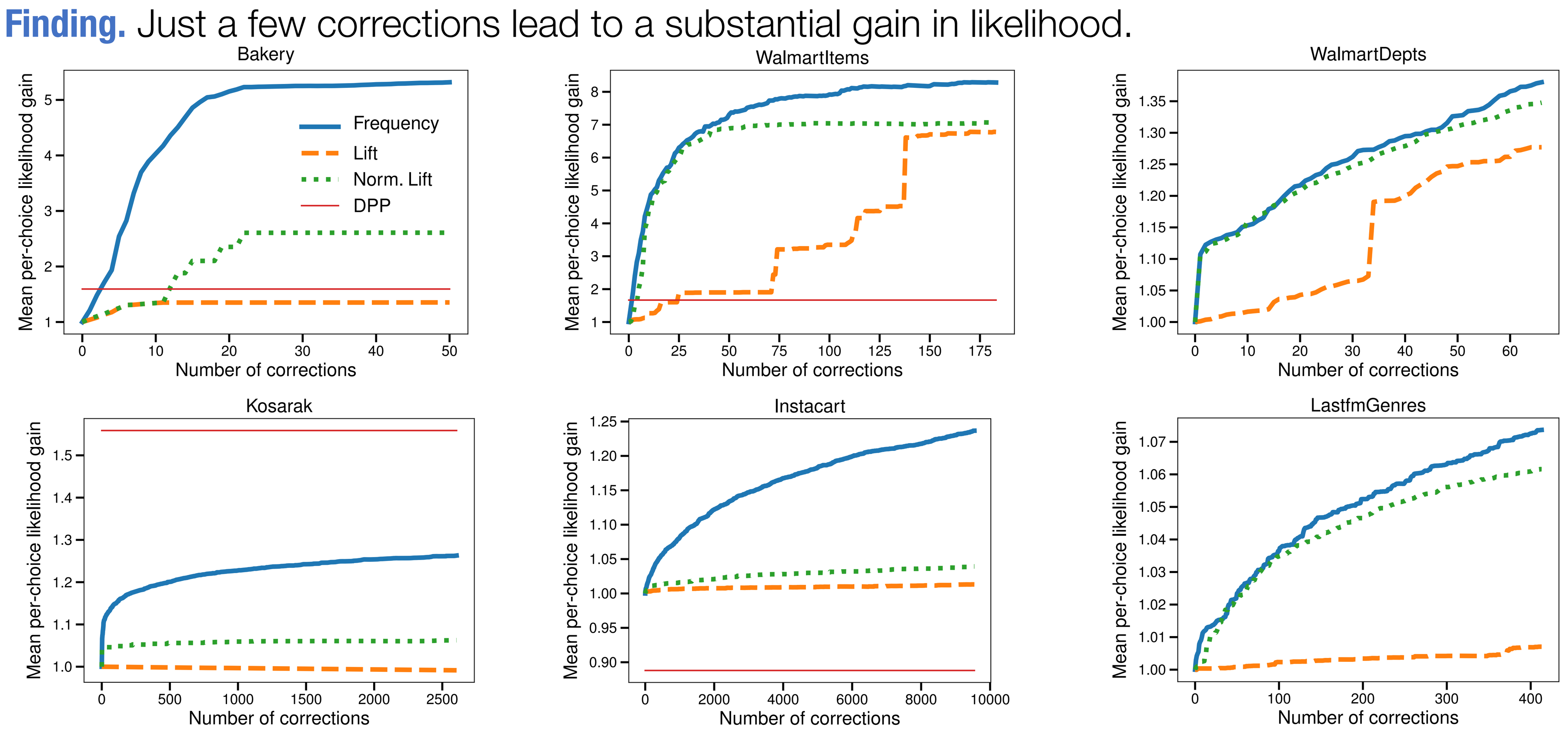
Dataset	# items	# choices	z_1	z_2	z_3	z_4	z_5
Bakery	50	67,488	0.05	0.20	0.37	0.25	0.13
WalmartItems	183	16,698	0.51	0.45	0.03	0.01	0.00
WalmartDepts	66	119,526	0.31	0.29	0.17	0.13	0.10
Kosarak	2,605	505,217	0.27	0.30	0.23	0.14	0.07
Instacart	9,544	806,662	0.19	0.21	0.21	0.21	0.19
LastfmGenres	413	643,982	0.52	0.21	0.12	0.08	0.06

Learning model parameters.

Theorem. Given a budget constraint on the number of special subsets (size of *H*), it is NP-hard to find the set will maximize likelihood (and it is also not a submodular optimization problem).

Theorem. Given H, there is a *closed form* for the model parameters that maximize likelihood.
Let $p^D_{ij} = N_{ij} / \sum_{\{k,l\}} N_{kl}$ be the empirical prob. of observing set $\{i,j\}$ in the data.
Then the MLE is:
(i) the p_i's are proportional to the number of times item *i* is selected in any set $\{i,j\} \notin H$. In other words, $p_i \propto \sum_{j:\{i,j\} \notin H} N_{ij}$;
(ii) $\gamma = (1 - \sum_{\{i,j\} \in H} p^D_{ij}) / (\sum_{\{i,j\} \notin H} p_i p_j)$;
(iii) given p & γ, q is set to match the empirical distribution of $\{i,j\}$: $\gamma p_i p_j + q_{ij} = p^D_{ij}$.

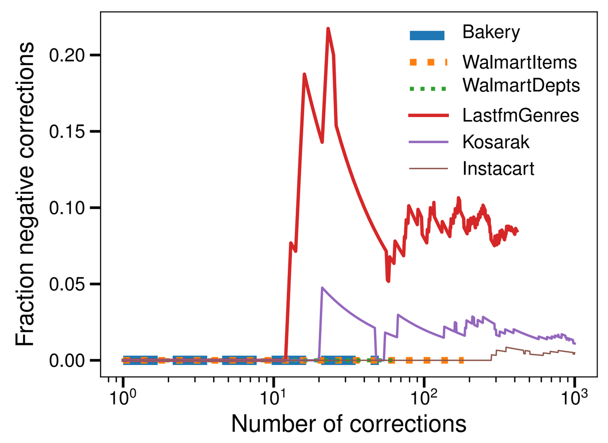
Algorithm. Use heuristic to find H, then use theorem to set model parameters.
Finding. Just a few corrections lead to a substantial gain in likelihood.



In practice, most correction probabilities q_{ij} are positive.

In these cases, the model has a different interpretation as a mixture of two multinomial logits.

1. With probability $a = \sum_{\{i,j\} \in H} q_{ij}$ follow the "full model" restricted to H.
2. With probability $1 - a$, follow the "separable model".



LastfmGenres dataset

Most positive
{indie, indie}
{rock, indie}
{hip_hop, hip_hop}
{indie, indie, indie}
{rock, rock, rock}
Most negative
{indie, metal}
{indie, progressive_metal}
{rock, rock, electronic}
{indie, industrial}
{metal, electronic}

Findings with "variable choice datasets"

Variable choice datasets: the set of available alternatives may be different for every subset choice.

Two datasets from YOOCHOOSE (Yc)

- **YcItems.** Subsets of items purchased from those viewed in a browsing session on an e-commerce web site
- **YcCats.** Subsets of item categories purchased from those viewed in a browsing session on an e-commerce web site

The z_k are the faction of choices that are size-*k* sets.

	YcItems	YcCats
# items	2,975	20
# choices	156,039	134,057
z_1	0.16	0.26
z_2	0.20	0.31
z_3	0.23	0.23
z_4	0.22	0.12
z_5	0.18	0.08

Learning model parameters.

Observations. (i) Still NP-hard to find best *H;* (ii) No closed form; but…
(iii) Given *H*, likelihood is a concave with linear constraints ⟶ easy to learn.

Finding. Again, just a few corrections (small size of *H*) lead to a substantial gain in likelihood.