Decision-Dependent Distributionally Robust Optimization for Facility Location Problem with Customer Behavior Considerations

Yutian He[1], Beste Basciftci[2], Xian Yu[3]

[1]Department of Mathematics, University of Iowa; [2]Department of Business Analytics, University of Iowa; [3]Department of Integrated Systems Engineering, The Ohio State University

1 Motivation

This work addresses a facility location problem with customer behavior considerations. We develop a Decision-Dependent Distributionally Robust Optimization (DDDRO) model that captures customer behavior as a second-stage recourse. Specifically, we utilize a ranking-based choice model informed by machine learning to represent consumer preferences.

2 Problem Setting

DDDRO: $$\min\left\{c^T x + \max_{\mathbb{P}_f \in U(x)} \mathbb{E}[g(x,\tilde{f})] : x \in X \subseteq \{0,1\}^{|N|\times|H|}\right\}$$

where

$$g(x,\tilde{f}) = \min_{\pi, L\geq 0} \sum_{q\in Q}\sum_{i\in N}\sum_{h\in H} r_{ih}\tilde{f}_q p_q \pi_{qih} + \sum_{i\in N}\sum_{h\in H} l_{ih}L_{ih} + F(\pi)$$

$$\text{s.t. } \pi_q \in \Pi_q(x) \subseteq \{0,1\}^{|N|\times|H|},$$

$$\sum_{q\in Q} \tilde{f}_q p_q \pi_{qih} \leq C_{ih}x_{ih} + L_{ih},\ \forall i \in N,\ h\in H,\ q\in Q.$$

- N: Facility location set.
- H: Facility type set.
- Q: Customer set.
- $\tilde{f}_q$: Random demand.
- c: Cost of opening facilities.
- r_{ih}: Revenue per unit of demand served.
- $F(\pi)$: Penalty for unsatisfied customer preference.
- x: First-stage facility location decisions.
- π: Second-stage customer adoption rate decisions.
- L: Second-stage unsatisfied demand decisions.
- p_q: Average consumption of customer's capacity.
- l_{ih}: Penalty per unit for unmet demand.
- C_{ih}: Maximum capacity for each type facility.

Ambiguity set

$$U(x) = \left\{\theta : \theta \in \mathbb{R}_+^{|K|}, \sum_{k=1}^{K}\theta^k = 1, \left|\sum_{k=1}^{K}\theta^k f^k - \mu_q(x)\right| \leq \epsilon_q^\mu,\ \forall q\in Q\right\}.$$

- K: Support set.
- $\mu_q(x)$: Expected mean with robustness parameter ϵ_q^μ.

3 Customer Behavior

$\Pi_q(x) = \{\pi_q : \pi_{qih} = 1 \text{ if } (i,h) \in \operatorname{argmax}_{(i,h)\in G_q(x)}\{u_{qih}\}, \text{and } \pi_{qih} = 0 \text{ otherwise}, \forall i\in N,\ h\in H\}$

$G_q(x)$: available (i,h) pairs under given x.

For every trip q, using utilities $u_{qih}, \forall\, i\in N, h\in H$, we can obtain ranking based customer choices (see Section 5)

Preference Ranking

Rankings $\{\sigma^1, \sigma^2, \dots, \sigma^{|Q|}\}$ over $|N|\times|H|+1$ options indicating each facility and type pair and customers not selecting any service.

Each ranking $\sigma^q: \{(0,0), \dots, (|N|\times|H|)\} \to \{0,1,\dots,|N|\times|H|\}$ is a bijection that assigns each option to a rank, where (0,0) denotes the customer choosing no facility.

Then we have $\pi_q \in \Pi_q(x) \subseteq \{0,1\}^{|N|\times|H|}$.

$$\sum_{i\in N\cup\{0\}}\sum_{h\in H}\pi_{qih} = 1, \pi_{qih} \leq x_{ih}, \sum_{(i',h'):\sigma^q(i',h')>\sigma^q(i,h)} \pi_{qi'h'} \leq 1 - x_{ih},$$

$$\sum_{(i',h'):\sigma^q(i',h')>\sigma^q(0,0)} \pi_{qi'h'} = 0, \forall i\in N, q\in Q, h\in H$$

4 Reformulations

Dual form for the second-stage problem:

$$\min_{\beta_q^1,\beta_q^2\geq 0,\alpha}\left\{\alpha + \sum_{q\in Q}\beta_q^1(\mu_q(x)+\epsilon_q^\mu) - \sum_{q\in Q}\beta_q^2(\mu_q(x)-\epsilon_q^\mu) \mid \alpha + \sum_{q\in Q}(\beta_q^1 f_q^k - \beta_q^2 f_q^k) \geq g(x,f^k),\ \forall k\right\}$$

Here, $\alpha, \beta_q^1, \beta_q^2$ are the dual variables associate with ambiguity set.

Then, the DDDRO full model can be written as the following

$$\min\quad c^T x + \alpha + \sum_{q\in Q}\beta_q^1(\mu_q(x)+\epsilon_q^\mu) - \sum_{q\in Q}\beta_q^2(\mu_q(x)-\epsilon_q^\mu)$$

$$\text{s.t.}\quad \alpha + \sum_{q\in Q}(\beta_q^1 f_q^k - \beta_q^2 f_q^k) \geq -\sum_{q\in Q}\sum_{i\in N}\sum_{h\in H} r_{ih}f_q^k p_q \pi_{qih}^k + \sum_{i\in N}\sum_{h\in H} l_{ih}L_{ih}^k + F(\pi),\quad \forall k\in K$$

$$\sum_{i\in N_q}\sum_{h}\pi_{qih}^k = 1,\quad \forall q\in Q, k\in K$$

$$\pi_{qih}^k \leq x_{ih},\quad \forall i\in N, q\in Q, h\in H, k\in K$$

$$\sum_{(i',h'):\sigma^q(i',h')>\sigma^q(i,h)} \pi_{qi'h'}^k \leq 1 - x_{ih},\quad \forall i\in N, q\in Q, h\in H, k\in K$$

$$\sum_{(i',h'):\sigma^q(i',h')>\sigma^q(0,0)} \pi_{qi'h'}^k = 0,\quad \forall q\in Q, k\in K$$

$$\sum_{q\in Q} f_q^k p_q \pi_{qih}^k \leq C_{ih}x_{ih} + L_{ih}^k\quad \forall i\in N, h\in H, k\in K$$

$$\sum_{h\in H} x_{ih} \leq 1\quad \forall i\in N$$

$$x_{ih}\in\{0,1\}\quad \forall i\in N, h\in H$$

$$\pi_{qih}^k \in \{0,1\}, \forall i\in N, q\in Q, h\in H, k\in K$$

$$L_{ih}^k \geq 0, \forall i\in N, h\in H, k\in K$$

Solution Algorithm

Column-and-Constraint Generation (CCG) method

Algorithm 1 Column-and-Constraint Generation Method (CCG)

Input: scenario set K, maximum iteration T, tolerance $\epsilon > 0$.
Initialize the master problem with first-stage constraints.
for $t = 0, 1, \dots, T$ **do**
  $(x_t, \alpha_t, \beta^1_{q(t)}, \beta^2_{q(t)}) \leftarrow$ solving the master problem.
  Fix x_t and construct the second-stage solution using closed form.
  Compute $\text{best\_val}_t = \max_{k\in K}\left\{g(x_t, f^k) - \sum_{q\in Q}(\beta^1_{q(t)} - \beta^2_{q(t)})f_q^k\right\}$.
  Let k_t^* be the scenario attaining best_val_t.
  if $\alpha_t + \varepsilon \leq \text{best\_val}_t$ **then**
    Add $\alpha + \sum_{q\in Q}(\beta_q^1 f_q^{k_t^*} - \beta_q^2 f_q^{k_t^*}) \geq g(x, f^{k_t^*})$ to the master problem.
    Add the second-stage constraints for scenario k_t^* to the master problem.
  else
    Break
  end if
end for
Return x_t.

5 Multinomial Logit (MNL) Choice Model

Utility function: $u_{qih} = v_{qih} + \varepsilon_{qih}$ where $v_{qih} = \beta_1 d_{qih} + \beta_{2h}$

Here, d_{qih} is the distance between trip q and station (i,h), β_1, β_{2h} are parameters to be estimated. We set the error ε_{qih} i.i.d. disturbances following the Gumbel distribution.

Then under this assumption, the probability of customer q choosing station i with type h can be written as

$$P_{qih}(S_q) = \frac{\exp(v_{qih})}{\sum_{j\in S_q}\exp(v_{qj})}$$

Where $S_q \subset [|N|\times|H|]$ denotes the choice set available to individual q.

Maximum Likelihood Estimation

$$\max_{\beta_1,\beta_{2h}} \sum_{j=1}^{J}\log\left(\frac{\exp(\beta_1 d_{jih} + \beta_{2h})}{\sum_{(i',h')\in[|N|\times|H|]}\exp(\beta_1 d_{ji'h'} + \beta_{2h})}\right)$$

J: number of data point

We use the Newton's method (e.g. SLSQP) to solve this problem.

Once we have β_1, β_2, we can finally define the ranking using utility functions by

$$\mathbb{E}[u_{qih}] > \mathbb{E}[u_{qi'h'}] \text{ if and only if } \sigma^q(i,h) < \sigma^q(i',h')$$

What happens if we consider customer behavior uncertainty (BU) ?

DDDRO+BU: $$\min\left\{c^T x + \max_{\mathbb{P}_f\in U(x)}\max_{m\in\mathcal{M}}\mathbb{E}[g(x,\tilde{f},\tilde{m})] : x\in X\subseteq\{0,1\}^{|N|\times|H|}\right\}$$

Here, M can be used to indicate different rankings, based on utility values.

6 Computational Study

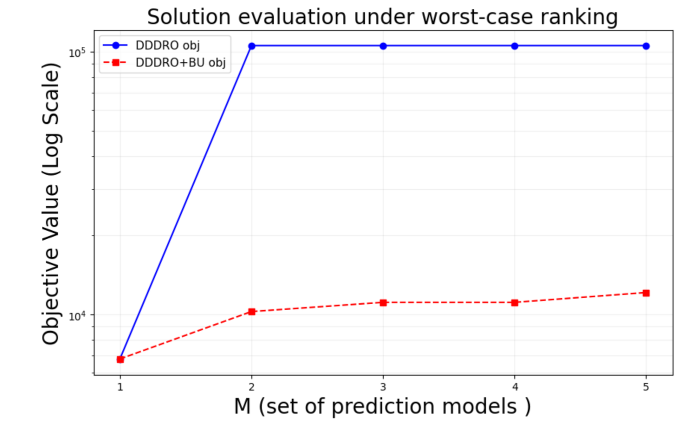
Computational performance under synthetic data, with M=1

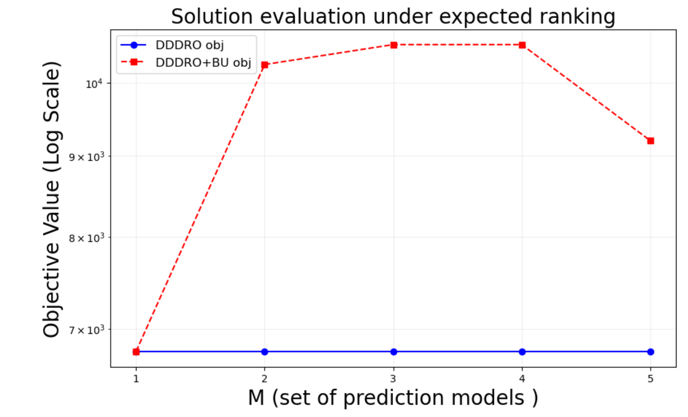
$\|K\|, \|Q\|, \|N\|, \|H\|$	Gurobi	CP		CCG		Speed up	
	RT (s)	RT (s)	Iter	RT (s)	Iter	Gurobi/CCG	CP/CCG
20, 10, 10, 2	13.49	39.37	4	2.14	3	6.30	18.40
20, 20, 20, 2	89.09	177.77	3	10.60	3	8.40	16.77
50, 10, 10, 2	34.75	99.02	4	2.06	3	16.87	48.07
50, 20, 20, 2	217.49	430.06	3	10.70	3	20.33	40.19
100, 10, 10, 2	71.80	190.81	4	2.18	3	32.94	87.53
100, 20, 20, 2	622.30	862.87	3	11.07	3	56.22	77.95

- CP: Cutting-Plane; RT: Runtime; Iter: Iteration; Speed up: ratio of runtime between different methods.

Computational performance under California EV charging data

		$m=1$	$m=2$	$m=3$	$m=4$	$m=5$	Avg. over rankings
DRO	Average cost	2751.14	39277.57	29447.77	43725.42	20123.18	27065.02
	Std. dev.	14.71	308.43	194.74	386.40	179.49	216.75
	90%	2770.40	39678.49	29706.99	44228.66	20351.77	27347.26
DDDRO	Average cost	3021.04	35504.76	33922.39	5966.89	18081.13	19299.24
	Std. dev.	16.64	303.40	224.80	24.63	159.24	145.74
	90%	3042.58	39508.02	34221.82	5997.93	18286.68	19491.41
DDDRO+BU under $M=5$	Average cost	6504.11	7647.77	7641.29	7075.19	2513.21	6276.31
	Std. dev.	26.99	31.98	32.55	27.46	20.54	27.90
	90%	6539.02	7689.91	7681.93	7109.63	2538.92	6311.88





- **The CCG algorithm consistently achieves the fastest convergence, outperforming both Reformulation and CP by several orders of magnitude as problem size increases.**
- **Under BU with different customer rankings, DDDRO+BU obtains the most stable results followed by DDDRO and then DRO.**
- **The first graph shows that DDDRO+BU achieves munch better solutions than DDDRO under the worst-case ranking in the BU uncertainty set.**
- **The second graph shows that DDDRO+ BU is more conservative than DDDRO under the baseline ranking model, but the gap is much smaller than the worst-case setting.**

7 References

[1] Beste Basciftci, Shabbir Ahmed, and Siqian Shen. Distributionally robust facility location problem under decision-dependent stochastic demand. *European Journal of Operational Research*, 292(2):548–561, 2021.

[2] Dimitris Bertsimas and Velibor V. Mišić. Exact first-choice product line optimization. *Operations Research*, 67(3):651–670, 2019.