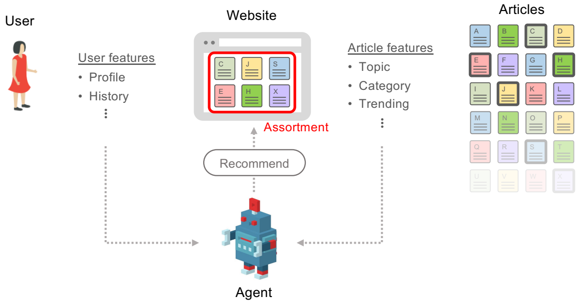
Nearly Minimax Optimal Regret for Multinomial Logistic Bandit

(NeurIPS 2024)

Joongkyu Lee & Min-hwan Oh

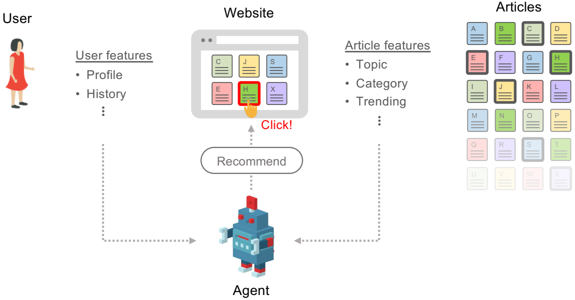
Seoul National University

Sequential Assortment Selection Problem



- Agent recommends an assortment (a set of items)
- User chooses one item from offered multiple options

Sequential Assortment Selection Problem



- Agent recommends an assortment (a set of items)
- User chooses one item from offered multiple options

Sequential Assortment Selection Problem

- For every round $t = 1, \ldots, T$:
  1. Observe contexts $x_{ti} \in \mathbb{R}^d$ and rewards $r_{ti} \in [0, 1]$ for every item $i \in [N]$

Sequential Assortment Selection Problem

- For every round $t = 1, \ldots, T$:
  1. Observe contexts $x_{ti} \in \mathbb{R}^d$ and rewards $r_{ti} \in [0, 1]$ for every item $i \in [N]$
  2. Offer an assortment $S_t = \{i_1, \ldots, i_m\}$ such that $m \leq K$

Sequential Assortment Selection Problem

- For every round $t = 1, \ldots, T$:
  1. Observe contexts $x_{ti} \in \mathbb{R}^d$ and rewards $r_{ti} \in [0, 1]$ for every item $i \in [N]$
  2. Offer an assortment $S_t = \{i_1, \ldots, i_m\}$ such that $m \leq K$
  3. Observe the user click decision $c_t \in S_t \cup \{0\}$ ("0": outside option)

Multinomial Logit (MNL) Choice Model (McFadden, 1977)

- Probability of choosing any item i in assortment S_t:

$$p_t(i|S_t, \mathbf{w}^\star) := \frac{\exp(x_{ti}^\top \mathbf{w}^\star)}{1+\sum_{j\in S_t}\exp(x_{tj}^\top \mathbf{w}^\star)}$$

Here, $\mathbf{w}^\star \in \mathbb{R}^d$ is an *unknown* parameter

Multinomial Logit (MNL) Choice Model (McFadden, 1977)

- Probability of choosing any item i in assortment S_t:

$$p_t(i|S_t, \mathbf{w}^\star) := \frac{\exp(x_{ti}^\top \mathbf{w}^\star)}{1+\sum_{j\in S_t}\exp(x_{tj}^\top \mathbf{w}^\star)}$$

Here, $\mathbf{w}^\star \in \mathbb{R}^d$ is an ***unknown*** parameter

- Expected revenue of the assortment S:

$$R_t(S, \mathbf{w}^\star) := \sum_{i\in S} p_t(i|S, \mathbf{w}^\star) r_{ti} = \frac{\sum_{i\in S}\exp(x_{ti}^\top \mathbf{w}^\star) r_{ti}}{1+\sum_{j\in S}\exp(x_{tj}^\top \mathbf{w}^\star)}$$

Multinomial Logit (MNL) Choice Model (McFadden, 1977)

- Probability of choosing any item i in assortment S_t:

$$p_t(i|S_t, \mathbf{w}^\star) := \frac{\exp(x_{ti}^\top \mathbf{w}^\star)}{1+\sum_{j \in S_t} \exp(x_{tj}^\top \mathbf{w}^\star)}$$

  Here, $\mathbf{w}^\star \in \mathbb{R}^d$ is an *unknown* parameter

- Expected revenue of the assortment S:

$$R_t(S, \mathbf{w}^\star) := \sum_{i \in S} p_t(i|S, \mathbf{w}^\star) r_{ti} = \frac{\sum_{i \in S} \exp(x_{ti}^\top \mathbf{w}^\star) r_{ti}}{1+\sum_{j \in S} \exp(x_{tj}^\top \mathbf{w}^\star)}$$

- Optimal assortment: $S_t^\star = \arg\max_{S \in \mathcal{S}} R_t(S, \mathbf{w}^\star)$

Multinomial Logit (MNL) Choice Model (McFadden, 1977)

- Probability of choosing any item i in assortment S_t:

$$p_t(i|S_t, \mathbf{w}^\star) := \frac{\exp(x_{ti}^\top \mathbf{w}^\star)}{1+\sum_{j\in S_t}\exp(x_{tj}^\top \mathbf{w}^\star)}$$

  Here, $\mathbf{w}^\star \in \mathbb{R}^d$ is an *unknown* parameter

- Expected revenue of the assortment S:

$$R_t(S, \mathbf{w}^\star) := \sum_{i\in S} p_t(i|S, \mathbf{w}^\star) r_{ti} = \frac{\sum_{i\in S}\exp(x_{ti}^\top \mathbf{w}^\star) r_{ti}}{1+\sum_{j\in S}\exp(x_{tj}^\top \mathbf{w}^\star)}$$

- Optimal assortment: $S_t^\star = \arg\max_{S\in\mathcal{S}} R_t(S, \mathbf{w}^\star)$
- Goal: Minimize $\mathbf{Reg}_T(\mathbf{w}^\star) = \sum_{t=1}^{T} R_t(S_t^\star, \mathbf{w}^\star) - R_t(S_t, \mathbf{w}^\star)$

Definitions

- **Uniform reward**: All items have the same reward (WLOG let $r_{ti} = 1$).
- **Non-uniform reward**: At every round t, reward r_{ti} for each item i is given arbitrarily.

Definitions

- **Uniform reward**: All items have the same reward (WLOG let $r_{ti} = 1$).
- **Non-uniform reward**: At every round t, reward r_{ti} for each item i is given arbitrarily.

- **Problem-dependent constant** κ:

$$\kappa := \min_{t \in [T]} \min_{S \in \mathcal{S}} \min_{\mathbf{w} \in \mathcal{W}} p_t(i|S, \mathbf{w}) p_t(0|S, \mathbf{w}),$$

where $\mathcal{W} := \{\mathbf{w} \in \mathbb{R}^d \mid \|\mathbf{w}\|_2 \leq 1\}$. Note that $1/\kappa = \mathcal{O}(K^2)$.

Previous Works

Table. T: total rounds, d: feature dimension, K: maximum assortment size, $1/\kappa = \mathcal{O}(K^2)$.

		Regret	Rewards	Comput. per Round
Lower Bound	Chen et al. (2020)	$\Omega(\frac{1}{K} d\sqrt{T})$	Uniform	–
Upper Bound	Oh and Iyengar (2019)	$\widetilde{\mathcal{O}}(\frac{1}{\kappa} d^{3/2}\sqrt{T})$	Non-uniform	$\mathcal{O}(t)$
	Chen et al. (2020)	$\widetilde{\mathcal{O}}(d\sqrt{T})$	Non-uniform	Intractable
	Oh and Iyengar (2021)	$\widetilde{\mathcal{O}}(\frac{1}{\kappa} d\sqrt{T})$	Non-uniform	$\mathcal{O}(t)$
	Perivier and Goyal (2022)	$\widetilde{\mathcal{O}}(d\sqrt{KT})$	Uniform	Intractable

1. No minimax result!

Previous Works

Table. T: total rounds, d: feature dimension, K: maximum assortment size, $1/\kappa = \mathcal{O}(K^2)$.

		Regret	Rewards	Comput. per Round
Lower Bound	Chen et al. (2020)	$\Omega(\frac{1}{K} d\sqrt{T})$	Uniform	–
Upper Bound	Oh and Iyengar (2019)	$\widetilde{\mathcal{O}}(\frac{1}{\kappa} d^{3/2}\sqrt{T})$	Non-uniform	$\mathcal{O}(t)$
	Chen et al. (2020)	$\widetilde{\mathcal{O}}(d\sqrt{T})$	Non-uniform	Intractable
	Oh and Iyengar (2021)	$\widetilde{\mathcal{O}}(\frac{1}{\kappa} d\sqrt{T})$	Non-uniform	$\mathcal{O}(t)$
	Perivier and Goyal (2022)	$\widetilde{\mathcal{O}}(d\sqrt{KT})$	Uniform	Intractable

1. No minimax result!
2. No lower bound under non-uniform rewards

Previous Works

Table. T: total rounds, d: feature dimension, K: maximum assortment size, $1/\kappa = \mathcal{O}(K^2)$.

		Regret	Rewards	Comput. per Round
Lower Bound	Chen et al. (2020)	$\Omega(\frac{1}{K} d\sqrt{T})$	Uniform	–
Upper Bound	Oh and Iyengar (2019)	$\widetilde{\mathcal{O}}(\frac{1}{\kappa} d^{3/2}\sqrt{T})$	Non-uniform	$\mathcal{O}(t)$
	Chen et al. (2020)	$\widetilde{\mathcal{O}}(d\sqrt{T})$	Non-uniform	Intractable
	Oh and Iyengar (2021)	$\widetilde{\mathcal{O}}(\frac{1}{\kappa} d\sqrt{T})$	Non-uniform	$\mathcal{O}(t)$
	Perivier and Goyal (2022)	$\widetilde{\mathcal{O}}(d\sqrt{KT})$	Uniform	Intractable

1. No minimax result!
2. No lower bound under non-uniform rewards
3. No computationally efficient algorithm

Main Contributions

Table. T: total rounds, d: feature dimension, K: maximum assortment size, $1/\kappa = \mathcal{O}(K^2)$.

		Regret	Rewards	Comput. per Round
Lower Bound	Chen et al. (2020)	$\Omega(\frac{1}{K} d\sqrt{T})$	Uniform	–
	This work	$\Omega(\frac{1}{\sqrt{K}} d\sqrt{T})$	Uniform	–
Upper Bound	Oh and Iyengar (2019)	$\widetilde{\mathcal{O}}(\frac{1}{\kappa} d^{3/2}\sqrt{T})$	Non-uniform	$\mathcal{O}(t)$
	Chen et al. (2020)	$\widetilde{\mathcal{O}}(d\sqrt{T})$	Non-uniform	Intractable
	Oh and Iyengar (2021)	$\widetilde{\mathcal{O}}(\frac{1}{\kappa} d\sqrt{T})$	Non-uniform	$\mathcal{O}(t)$
	Perivier and Goyal (2022)	$\widetilde{\mathcal{O}}(d\sqrt{KT})$	Uniform	Intractable
	This work	$\widetilde{\mathcal{O}}(\frac{1}{\sqrt{K}} d\sqrt{T})$	Uniform	$\mathcal{O}(1)$

1. Close gap between upper and lower bounds:
   - **Uniform** rewards: $K \uparrow \implies \mathbf{Reg}_T \downarrow$

Main Contributions

Table. T: total rounds, d: feature dimension, K: maximum assortment size, $1/\kappa = \mathcal{O}(K^2)$.

		Regret	Rewards	Comput. per Round
Lower Bound	Chen et al. (2020)	$\Omega(\frac{1}{K} d\sqrt{T})$	Uniform	–
	This work	$\Omega(\frac{1}{\sqrt{K}} d\sqrt{T})$	Uniform	–
	This work	$\Omega(d\sqrt{T})$	Non-uniform	–
Upper Bound	Oh and Iyengar (2019)	$\widetilde{\mathcal{O}}(\frac{1}{\kappa} d^{3/2}\sqrt{T})$	Non-uniform	$\mathcal{O}(t)$
	Chen et al. (2020)	$\widetilde{\mathcal{O}}(d\sqrt{T})$	Non-uniform	Intractable
	Oh and Iyengar (2021)	$\widetilde{\mathcal{O}}(\frac{1}{\kappa} d\sqrt{T})$	Non-uniform	$\mathcal{O}(t)$
	Perivier and Goyal (2022)	$\widetilde{\mathcal{O}}(d\sqrt{KT})$	Uniform	Intractable
	This work	$\widetilde{\mathcal{O}}(\frac{1}{\sqrt{K}} d\sqrt{T})$	Uniform	$\mathcal{O}(1)$
	This work	$\widetilde{\mathcal{O}}(d\sqrt{T})$	Non-uniform	$\mathcal{O}(1)$

1. Close gap between upper and lower bounds:
   - **Uniform** rewards: $K \uparrow \implies \mathbf{Reg}_T \downarrow$
   - **Non-uniform** rewards: $\mathbf{Reg}_T$ is NOT affected by K

Main Contributions

Table. T: total rounds, d: feature dimension, K: maximum assortment size, $1/\kappa = \mathcal{O}(K^2)$.

		Regret	Rewards	Comput. per Round
Lower Bound	Chen et al. (2020)	$\Omega(\frac{1}{K} d\sqrt{T})$	Uniform	–
	This work	$\Omega(\frac{1}{\sqrt{K}} d\sqrt{T})$	Uniform	–
	This work	$\Omega(d\sqrt{T})$	Non-uniform	–
Upper Bound	Oh and Iyengar (2019)	$\widetilde{\mathcal{O}}(\frac{1}{\kappa} d^{3/2}\sqrt{T})$	Non-uniform	$\mathcal{O}(t)$
	Chen et al. (2020)	$\widetilde{\mathcal{O}}(d\sqrt{T})$	Non-uniform	Intractable
	Oh and Iyengar (2021)	$\widetilde{\mathcal{O}}(\frac{1}{\kappa} d\sqrt{T})$	Non-uniform	$\mathcal{O}(t)$
	Perivier and Goyal (2022)	$\widetilde{\mathcal{O}}(d\sqrt{KT})$	Uniform	Intractable
	This work	$\widetilde{\mathcal{O}}(\frac{1}{\sqrt{K}} d\sqrt{T})$	Uniform	$\mathcal{O}(1)$
	This work	$\widetilde{\mathcal{O}}(d\sqrt{T})$	Non-uniform	$\mathcal{O}(1)$

1. Close gap between upper and lower bounds:
   - **Uniform** rewards: $K \uparrow \implies \mathbf{Reg}_T \downarrow$
   - **Non-uniform** rewards: $\mathbf{Reg}_T$ is NOT affected by K
2. First lower bound for non-uniform rewards

Main Contributions

Table. T: total rounds, d: feature dimension, K: maximum assortment size, $1/\kappa = \mathcal{O}(K^2)$.

		Regret	Rewards	Comput. per Round
Lower Bound	Chen et al. (2020)	$\Omega(\frac{1}{K} d\sqrt{T})$	Uniform	–
	This work	$\Omega(\frac{1}{\sqrt{K}} d\sqrt{T})$	Uniform	–
	This work	$\Omega(d\sqrt{T})$	Non-uniform	–
Upper Bound	Oh and Iyengar (2019)	$\widetilde{\mathcal{O}}(\frac{1}{\kappa} d^{3/2}\sqrt{T})$	Non-uniform	$\mathcal{O}(t)$
	Chen et al. (2020)	$\widetilde{\mathcal{O}}(d\sqrt{T})$	Non-uniform	Intractable
	Oh and Iyengar (2021)	$\widetilde{\mathcal{O}}(\frac{1}{\kappa} d\sqrt{T})$	Non-uniform	$\mathcal{O}(t)$
	Perivier and Goyal (2022)	$\widetilde{\mathcal{O}}(d\sqrt{KT})$	Uniform	Intractable
	This work	$\widetilde{\mathcal{O}}(\frac{1}{\sqrt{K}} d\sqrt{T})$	Uniform	$\mathcal{O}(1)$
	This work	$\widetilde{\mathcal{O}}(d\sqrt{T})$	Non-uniform	$\mathcal{O}(1)$

1. Close gap between upper and lower bounds:
   - **Uniform** rewards: $K \uparrow \implies \mathbf{Reg}_T \downarrow$
   - **Non-uniform** rewards: $\mathbf{Reg}_T$ is NOT affected by K
2. First lower bound for non-uniform rewards
3. Propose computationally efficient, nearly minimax optimal algorithm

References I

Chen, X., Wang, Y., and Zhou, Y. (2020). Dynamic assortment optimization with changing contextual information. The Journal of Machine Learning Research, 21(1):8918–8961.

McFadden, D. (1977). Modelling the choice of residential location.

Oh, M.-h. and Iyengar, G. (2019). Thompson sampling for multinomial logit contextual bandits. Advances in Neural Information Processing Systems, 32.

Oh, M.-h. and Iyengar, G. (2021). Multinomial logit contextual bandits: Provable optimality and practicality. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 35, pages 9205–9213.

Perivier, N. and Goyal, V. (2022). Dynamic pricing and assortment under a contextual mnl demand. Advances in Neural Information Processing Systems, 35:3461–3474.