Contextual Decision-Making with Knapsacks Beyond the Worst Case

Zhaohua Chen[1], Rui Ai[2], Mingwei Yang[3], Yuqi Pan[4],
Chang Wang[5], and Xiaotie Deng[1]

NeurIPS 2024

[1]Peking University, [2]MIT, [3]Stanford University, [4]Harvard University, [5]Northwestern University

Model

Goal: maximize total reward under the initial resource constraint!

A warehouse

Resource inventory: $\boldsymbol{B} = \boldsymbol{\rho} \cdot T$, n resources

Stochastic request $\theta_t \in [k]$:
Deliver goods to places

Action $a_t \in A = [m] \cup \{0\}$:
Choose a way to deliver (A^+), or reject

Stochastic external factor γ_t:
Affects the resource cost and reward

Consumes resources $\boldsymbol{c}(\theta_t, a_t, \gamma_t)$, receives reward $r(\theta_t, a_t, \gamma_t)$.

Model

NEURAL INFORMATION PROCESSING SYSTEMS

Stochastic request $\theta_t \in [k]$:
Deliver goods to places

Stochastic external factor γ_t:
Affects the resource cost and reward

- Unknown distribution for request θ and external factor γ!
- Information model:
  - [Full feedback.] Always learns γ_t after the round.
  - [Partial feedback.] Learns γ_t only when a_t is not a reject.

Previous Methods

- Best-policy method [ADL16, …]
  - Pick the best policy in a UCB manner
  - An $O\left(\sqrt{mT \log nT}\right)$ regret
- Dual-update method [SSF23, …]
  - Lagrangian-based control with regression oracle
  - An $O\left(\sqrt{nT \log nT}\right)$ regret
- Work with bandit feedback, but under certain assumptions
- Our method: re-solving-based

- m: number of actions
- n: number of resources
- k: size of request space

Benchmarks

- Online optimum (V^{ON}) is hard to compute and analyze!
- Fluid optimum (V^{FL}): maximum expected reward under the expected resource constraint. Often used as the benchmark. $V^{\mathrm{FL}} \geq V^{\mathrm{ON}}$.

$$V^{\mathrm{FL}} := T \cdot \max_{\phi} \mathbb{E}_{\theta}\left[\sum_{a \in A^{+}} \mathbb{E}_{\gamma}[r(\theta, a, \gamma)]\phi(\theta, a)\right],$$

$$\text{s.t.} \quad \mathbb{E}_{\theta}\left[\sum_{a \in A^{+}} \mathbb{E}_{\gamma}[\boldsymbol{c}(\theta, a, \gamma)]\phi(\theta, a)\right] \leq \boldsymbol{\rho};$$

$$\sum_{a \in A^{+}} \phi(\theta, a) \leq 1, \forall \theta; \quad \phi(\theta, a) \geq 0, \forall \theta, a.$$

Theorem. When V^{FL} has a unique and degenerate optimal solution, $V^{\mathrm{FL}} - V^{\mathrm{ON}} = \Omega(\sqrt{T})$.

The Re-Solving Heuristic

NEURAL INFORMATION PROCESSING SYSTEMS

- In each round t:
  - Solve the approximated fluid optimum $\hat{J}(\boldsymbol{\rho}_t)$ with respect to the remaining average resource constraint $\boldsymbol{\rho}_t$ and estimated distributions of the request and external factor, and obtain $\hat{\phi}_t$.
  - Observe θ_t, and act according to the distribution $\hat{\phi}_t(\theta_t, \cdot)$.
  - Update estimated distributions according to the observation.

$$\hat{J}(\boldsymbol{\rho}_t) \coloneqq T \cdot \max_{\hat{\phi}_t} \mathbb{E}_\theta \left[\sum_{a \in A^+} \mathbb{E}_\gamma [r(\theta, a, \gamma)] \hat{\phi}_t(\theta, a) \right],$$

$$\text{s.t.} \quad \mathbb{E}_\theta \left[\sum_{a \in A^+} \mathbb{E}_\gamma [\boldsymbol{c}(\theta, a, \gamma)] \hat{\phi}_t(\theta, a) \right] \le \boldsymbol{\rho}_t; \quad \sum_{a \in A^+} \hat{\phi}_t(\theta, a) \le 1, \forall \theta; \quad \hat{\phi}_t(\theta, a) \ge 0, \forall \theta, a.$$

The Re-Solving Heuristic -- Guarantee

NEURAL INFORMATION PROCESSING SYSTEMS

- ***Fluid program has a unique and non-degenerate solution***

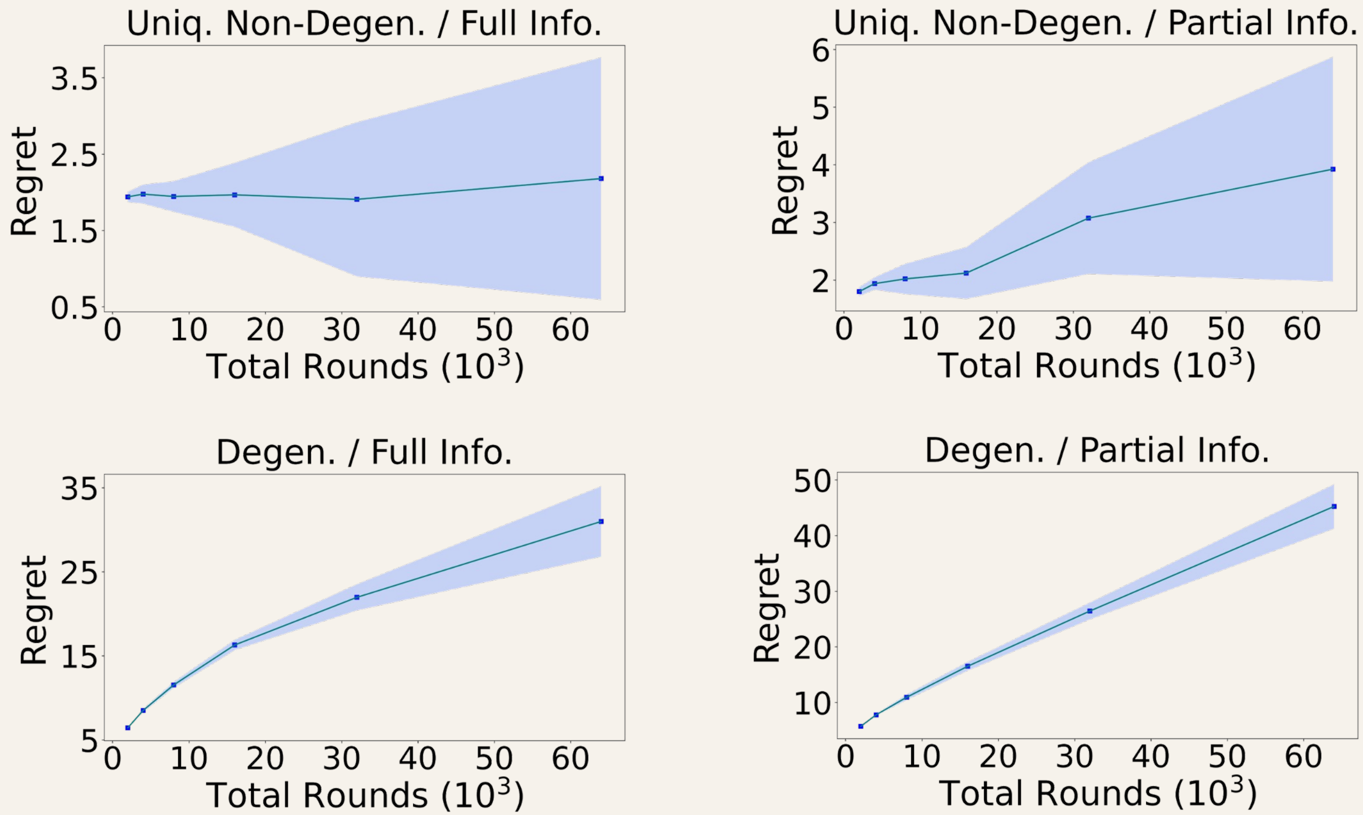
- **Stability factor** D: L_∞ distance of the fluid program to any program with non-unique or degenerate solution(s).
  - Full feedback: $O\left(\frac{n^2+k}{D^2}\right)$ gap to the fluid optimum.
  - Partial feedback: $O\left(\frac{n^2+k+\log T}{D^2}\right)$ gap to the fluid optimum.

The Re-Solving Heuristic -- Guarantee

- ***No assumption on the fluid program***
  - Full feedback: $O\left(k\sqrt{T\log T}+n\right)$ gap to the fluid optimum.
  - Partial feedback: $O\left(k\sqrt{T}\log T+n\right)$ gap to the fluid optimum.

- ***Can be generalized to continuous randomness***
  - With non-parameterized estimation methods

Numerical Validations



NEURAL INFORMATION PROCESSING SYSTEMS

Thank you!