Small Total-Cost Constraints in Contextual Bandits with Knapsacks [CBwK], with Application to Fairness

Evgenii Chzhen – Christophe Giraud – Gilles Stoltz

Zhen Li

Neurips 2023

CBwK framework – Novelty is $T\mathbf{B}$ with components of order $\sqrt{T}$

Known: finite $\mathcal{A}$, rounds T, average costs $\mathbf{B} \in [0,1]^d$

Unknowns:

Context distribution ν on $\mathcal{X}$
Scalar mean-payoff function $r : \mathcal{X} \times \mathcal{A} \to [0,1]$
Vector-valued mean-cost function $\mathbf{c} : \mathcal{X} \times \mathcal{A} \to [-1,1]^d$

For rounds $t = 1, 2, \ldots, T$:

Observe context $\mathbf{x}_t \sim \nu$, and pick $a_t \in \mathcal{A}$
Get payoff r_t and costs $\mathbf{c}_t$ with cond. exp. $r(\mathbf{x}_t, a_t)$ and $\mathbf{c}(\mathbf{x}_t, a_t)$

Goals (cf. fairness costs: $T\mathbf{B}$ as small as possible, possibly $\sqrt{T}$)

$$\text{Ensure} \quad \sum_{t=1}^{T} c_t \leqslant T\,\mathbf{B} \quad \text{a.s. while maximizing} \quad \sum_{t=1}^{T} r_t$$

First reference for CBwK: Badanidiyuru, Langford, Slivkins [2014]

State of the art = $T\mathbf{B}$ at best $T^{3/4}$: Agrawal and Devanur [2016], Han et al. [2022]

Regret: Minimize $\qquad R_T = T\,\mathrm{opt}(r, \mathbf{c}, \mathbf{B}) - \sum_{t=1}^{T} r_t \qquad$ where

$$\begin{aligned}
&\mathrm{opt}(r, \mathbf{c}, \mathbf{B}) \\
&= \sup_{\pi:\mathcal{X}\to\mathcal{P}(\mathcal{A})} \left\{ \mathbb{E}_{\mathbf{X}\sim\nu}\left[\sum_{a\in\mathcal{A}} r(\mathbf{X}, a)\,\pi_a(\mathbf{X})\right] : \;\; \mathbb{E}_{\mathbf{X}\sim\nu}\left[\sum_{a\in\mathcal{A}} \mathbf{c}(\mathbf{X}, a)\,\pi_a(\mathbf{X})\right] \leqslant \mathbf{B} \right\} \\
&= \sup_{\boldsymbol{\pi}:\mathcal{X}\to\mathcal{P}(\mathcal{A})} \; \inf_{\boldsymbol{\lambda}\geqslant\mathbf{0}} \; \mathbb{E}_{\mathbf{X}\sim\nu}\left[\sum_{a\in\mathcal{A}} r(\mathbf{X}, a)\,\pi_a(\mathbf{X}) + \left\langle \boldsymbol{\lambda},\, \mathbf{B} - \sum_{a\in\mathcal{A}} \mathbf{c}(\mathbf{X}, a)\,\pi_a(\mathbf{X}) \right\rangle \right] \\
&= \min_{\boldsymbol{\lambda}\geqslant\mathbf{0}} \; \mathbb{E}_{\mathbf{X}\sim\nu}\left[\max_{a\in\mathcal{A}}\left\{ r(\mathbf{X}, a) - \left\langle \mathbf{c}(\mathbf{X}, a) - \mathbf{B},\, \boldsymbol{\lambda} \right\rangle \right\}\right]
\end{aligned}$$

$\rightarrow$ Suffices to learn r and c, as well as $\lambda^\star$

Learn r and c: via structural assumptions; uniform bounds

Linear model: Agrawal and Devanur [2016], based on LinUCB from Abbasi-Yadkori et al. [2011]. Logistic model: Li and Stoltz [2022], based on Logistic-UCB1 from Faury et al. [2020].

Target: $\quad \mathrm{opt}(r, \mathbf{c}, \mathbf{B}) = \min_{\boldsymbol{\lambda} \geqslant \mathbf{0}} \mathbb{E}_{\mathbf{X}\sim\nu}\left[\max_{a\in\mathcal{A}}\left\{ r(\mathbf{X}, a) - \left\langle \mathbf{c}(\mathbf{X}, a) - \mathbf{B},\, \boldsymbol{\lambda} \right\rangle \right\}\right]$

→ Gradient descent on dual / best response for primal var.

Algorithm with fixed step size γ

For $t = 1, 2, \ldots, T$:

1. Play $a_t \in \arg\max_{a\in\mathcal{A}} \left\{ \widehat{r}_{t-1}(\mathbf{x}_t, a) - \left\langle \widehat{\mathbf{c}}_{t-1}(\mathbf{x}_t, a) - (\mathbf{B} - b\mathbf{1}),\, \boldsymbol{\lambda}_{t-1} \right\rangle \right\}$
2. Make gradient step $\boldsymbol{\lambda}_t = \Big(\boldsymbol{\lambda}_{t-1} + \gamma\big(\widehat{\mathbf{c}}_{t-1}(\mathbf{x}_t, a) - (\mathbf{B} - b\mathbf{1})\big)\Big)_+$
3. Update estimates $\widehat{r}_t$ and $\widehat{\mathbf{c}}_t$ of functions r and $\mathbf{c}$

Analysis

Cost margin Tb, of order $\big(1 + \|\lambda^\star\|\big)/\gamma$; adds $\|\lambda^\star\|\big(Tb + \sqrt{T}\big)$ to regret

→ Oracle choice $\big(1 + \|\lambda^\star\|\big)/\sqrt{T}$ for γ, leads to $\big(1 + \|\lambda^\star\|\big)\sqrt{T}$ regret

Reminder of the issue: oracle choice $(1 + \|\lambda^\star\|)/\sqrt{T}$ for γ

Typical bypass by estimating $\|\lambda^\star\|$ on $\sqrt{T}$ preliminary rounds (see, e.g.: Agrawal and Devanur [2016], Han et al. [2022]) leads to $\min \mathbf{B} \geqslant T^{-1/4}$

Theorem

Algorithm based on a careful doubling trick $\gamma_k = 2^k/\sqrt{T}$
W.h.p.: controls cumulative costs & regret of order $(1 + \|\lambda^\star\|)\sqrt{T}$
Only requires $\min \mathbf{B}$ *to be larger than* $1/\sqrt{T}$ *up to poly-log terms*

Note 1: if null-cost action, $\|\lambda^\star\| \leqslant \dfrac{2\,\mathrm{opt}(r, \mathbf{c}, \mathbf{B})}{\min \mathbf{B}}$ = usual bound

Note 2: explicit, closed-form bounds in the article

Note 3: fairness example from Chohlas-Wood et al. [2021]