Optimizing the coalition gain in Online Auctions with Greedy Structured Bandits

D. Baudry, H. Richard, M. Cherifa, C. Calauzènes, V. Perchet

March 14, 2025

   

- **Public (old) Auctions**
  1. **User** u arrives, with some features X_u (irrelevant for us)
  2. **DSP** (us) runs N campaigns, observe $v_{u,1}, v_{u,2}, \ldots, v_{u,N}$
  3. DSP **bids** $\max_{i \in [N]} v_{u,i}$
  4. **Competition** bids $v_{u,N+1}, \ldots, v_{u,N+p}$
  5. 2nd price auction. Winner $\arg\max v_{u,j}$, pays 2nd-highest

- **Private (future) Auctions**
  1. **User** u arrives, its features X_u are **not observed**
  2. **DSP** (us) **Only knows** $v_{??,1} \sim F_1, v_{??,2}, \ldots, v_{??,N} \sim F_N$
  3. DSP do not bid but **selects subset of compaigns** $\mathcal{N} \subset [N]$
  4. **Competition** bids $v_{u,n+1}, \ldots, v_{u,n+p}$
  5. Winner $\arg\max_{j \in \mathcal{N} \cup \{n+1,\ldots,n+p\}} v_{u,j}$, pays 2nd-highest

- Choosing a larger number of ads impacts the **outcome**:

  Increases the probability of winning

  Decreases the gain from winning

- Larger size also impacts the **observations**

  Increases the proba. of observing (a click or not)

  Decreases the observation quality (high variance)

$\Longrightarrow$ Tradeoff in choosing “coalition size”

- Model (new, future) **privacy constraints** in online advertising

- T ad slots sold sequentially through **second price auctions**.
  Highest bidder wins, pays second highest bid
- The DSP chooses $n_t \leq N$ campaigns that *participate*
- There are $p \in \mathbb{N}^*$ external competitors.
- All $N + p$ bidders' valuation are i.i.d. $v_{n,t} \sim F$ the **unknown** cdf
  Bidders bid truthfully their value, $b_{n,t} = v_{n,t}$
- DSP only observes the reward and value if the coalition wins.

The reward and regret

- If coalition chooses n bidders to participate, its reward is

$$r(n) := \mathbb{E}_{\mathbf{v}=(v_i)_{i\in[n+p]}\sim F^{\otimes n+p}}\left[(\mathbf{v}_{(1)} - \mathbf{v}_{(2)})\mathbb{1}\left\{\arg\max_{i\in[n+p]} v_i \in [n]\right\}\right]$$

where $\mathbf{v}_{(1)}$ and $\mathbf{v}_{(2)}$ are first and second maximum of $\mathbf{v}$.

- Sequence of choices $n_1, \ldots, n_T$ leads to regret

$$\mathcal{R}_T = \sum_{t\leq T} r(n^*) - r(n_t) \ , \quad \text{with} \quad n^* = \operatorname*{argmax}_{n\in[N]} r(n)$$

- Standard bandit algorithms $\mathcal{R}_T \leq \tilde{\mathcal{O}}(\min\{\frac{N\log(T)}{\Delta}, \sqrt{NT}\})$

$\Longrightarrow$ Leverage structure to **improve guarantees** ?

The estimation

Using order statistics properties, the reward function is satisfies,

$$r(n) = \underbrace{n \int_0^1 F^{p+n-1}(x) - F^{p+n}(x) \mathrm{d}x}_{n \text{ times a decreasing function with } n} \tag{1}$$

$\Longrightarrow$ $r(n)$ is usually unimodal (at least for lots of cdf F)!

$$r(n) = \underbrace{n\int_0^1 F^{p+n-1}(x) - F^{p+n}(x)\mathrm{d}x}_{\text{estimating } F^{n+p-1} \text{ and } F^{n+p} \text{ is sufficient to estimate } r(n)}$$

- n not fixed in advance!

  $\Longrightarrow$ Need an estimator for any power F^m.
- A sample of F^{n_t+p} gathered if auction t is won (the winning bid)
  - Combining samples from different F^{n_t+p} challenging
  - $\hat{F}^m = \left(\hat{F}^k\right)^{\frac{m}{k}}$ much simpler, if m and k **not too different**

- Past winning bids when $n_t = k$ $\overline{W_k} = (w_{k,1}, \dots, w_{k,m_k})$
- **Empirical cdf of** F^{k+p} : $\hat{F}_{k+p}(x) = \frac{1}{m_k} \sum_{j=1}^{m_k} \mathbb{1}\{w_{k,j} \leq x\}$
- Estimations
  - of powers $\tilde{F}_{k+p}^{n+p}(x) = \hat{F}_{k+p}^{\frac{n+p}{k+p}}(x)$
  - of reward function (n different estimators)

$$\widehat{r}_k(n) = n \int_0^1 \left(\widetilde{F}_{k+p}^{n+p-1}(x) - \widetilde{F}_{k+p}^{n+p}(x) \right) \mathrm{d}x$$

⚠ k and n should be **close enough**

$$F(x)^n - \widehat{F}_k(x)^{\frac{n}{k}} \approx \frac{n}{k} F_k(x)^{\frac{n}{k}} (F(x)^k - \widehat{F}_k(x)) \frac{1}{F(x)}$$

- $n \geq k$, error scales as n/k
- $n < k$, error scales with $1/F(x)$

Theorem (informal)

Fix $n \leq N$*, then for any* $k \in \mathcal{N}(n) := \left[\frac{n+p}{2} - p, \frac{3}{2}(n+p-1) - p\right]$*,*
with probability $1-\delta$*,*

$$|\hat{r}_k(n) - r(n)| \lesssim \sqrt{\frac{\log\left(\frac{nm_k}{\delta}\right)}{m_k}} + n\left(\frac{\log\left(\frac{nm_k}{\delta}\right)}{m_k}\right)^{\frac{n+p-1}{k+p}} .$$

- The n term becomes $L\log(n)$ if F L-Lipschitz
- Technical proof on **concentration ineq.**
- Can **estimate** $r(n)$ from any k in its neighborhood $\mathcal{N}(n)$

  the one with **the most samples** !

The algorithms

Idea: adaptation of OSUB (Combes and Proutière 2014).

Algorithm Local Greedy LG

Input: exploration parameter α, neighborhoods $\mathcal{N}(n)$

Play $n_1 = 1$ and observe $w \sim F^{1+p}$; ▷ Initialization

for $t \geq 2$ **do**

  Set $\ell_t = n_{t-1}$, compute $(\hat{r}_{\ell_t}(n))_{n \in \mathcal{V}(\ell_t)}$; ▷ Estimate from leader

  if $m_t := |\{s \in [t-1], n_s = \ell_t\}| \leq \alpha t$ **then**

    play $n_t = \ell_t$; ▷ Linear sampling

  else

    play $n_t \in \mathrm{argmax}_{n \in \mathcal{V}(\ell_t)} \hat{r}_{\ell_t}(n)$; ▷ Greedy play in $\mathcal{N}(\ell_t)$

  Observe $w \sim F^{n_t+p}$; ▷ Gather feedback

Theorem (informal)

Let $\Delta := \min_{n\in[N-1]} |r(n+1) - r(n)|$ *(worst local gap) and* $\Delta_n = r(n^*) - r(n)$*. The regret of* LG *is* ***bounded*** *and satisfies*

$$\mathcal{R}_T \leq \tilde{\mathcal{O}}_N\Big(\sum_{n\in[N]} \frac{\Delta_n}{\Delta^2}\Big)$$

- ✓ Works thanks to unimodality:

  there is a better decision in the neighborhood of the empirical best one in the direction of the optimal.

- ✗ The regret of LG depends on the **worst local gap**!

Greedy Grid = Local Greedy + Successive Elimination

Algorithm Greedy Grid

Input: Grid $\mathcal{S}$, confidence levels $(\delta_t)_{t\in\mathbb{N}}$, sampling parameter α

Play $n_1 = \min \mathcal{S}$ and observe $w \sim F^{n_1+p}$

for $t \geq 2$ *and* $n \in [N]$ **do**

- $\ell_n = \mathrm{argmax}_{k\in\mathcal{V}(n)}\, m_k(t)$; ▷ Elect leaders
- $L_n = \widehat{L}_{\ell_n}(n, \delta_t)$ and $U_n = \widehat{U}_{\ell_n}(n, \delta_t)$; ▷ Compute UCB and LCB
- $i_t^* = \mathrm{argmax}_{n\in[N]}\, L_n$; ▷ Compute best lower bound index
- $\mathcal{C}_t = \{a \in \mathcal{S}, U_s \geq L_{i_t^*}, \forall s \in [a, i_t^*]\}$; ▷ Remaining grid arms
- **if** $n_{t-1} \in B(i_t^*)$ *and* $m_{n_{t-1}} \leq \alpha t$ **then**
  - Play $n_t = n_{t-1}$ ▷ linear sampling
- **else** ▷ Play unif in grid or greedy
  - **If** $\mathcal{C}_t \neq \emptyset$: Round Robin on $\mathcal{C}_t$ **Else** play $\mathrm{argmax}_{n\in B(i_t^*)}\, \hat{r}_{\ell_n}(n)$
- Observe $w \sim F^{n_t+p}$

Theorem (informal)

Suppose that GG is tuned with confidence level $\delta_t = \frac{1}{N^2 t^3}$, *and* $\alpha = 1/4$. *Then, for any* $T \in \mathbb{N}$ *it holds that*

$$\mathcal{R}_T \leq \tilde{\mathcal{O}}\Big(\sum_{n \in \mathcal{B}^\star} \frac{1}{\Delta_n} + \sum_{k \in \mathcal{S}} \frac{1}{\Delta_k}\Big)$$

- $\mathcal{B}^\star$ is the bin of arm $n^\star$.

✓ **No** dependence on the worst **local gap** anymore!

✓ $\mathcal{R}_T \leq \mathcal{O}(\sqrt{(\log(N) + |\mathcal{B}^\star|)T}) = \mathcal{O}(\sqrt{(\log(N) + n^\star)T})$

A benchmark of `LG`, `GG`, `UCB`, `EXP3` and `OSUB` on synthetic data in terms of the expected regret $\mathcal{R}(T)$.

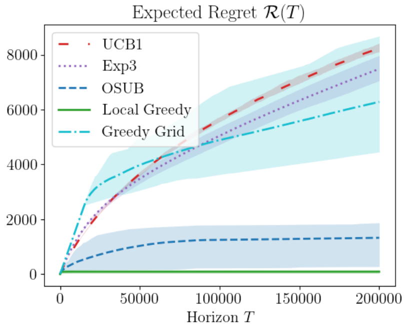

Figure: Performance of `LG` and `GG`, `OSUB`, `UCB` and `EXP3`, computed over 20 trajectories, with $\mathcal{B}(0.05)$, $N = 100$ and $p = 4$

Thank you