Finding the bandit in a graph: Sequential search-and-stop

Pierre Perrault[1,2], Vianney Perchet[2,3], Michal Valko[1]

[1]SequeL team, INRIA Lille [2]CMLA, ENS Paris-Saclay [3]Criteo Research

Background

Sequential search-and-stop problem

$\mathcal{G} = ([n], \mathcal{E})$ is a fixed DAG.

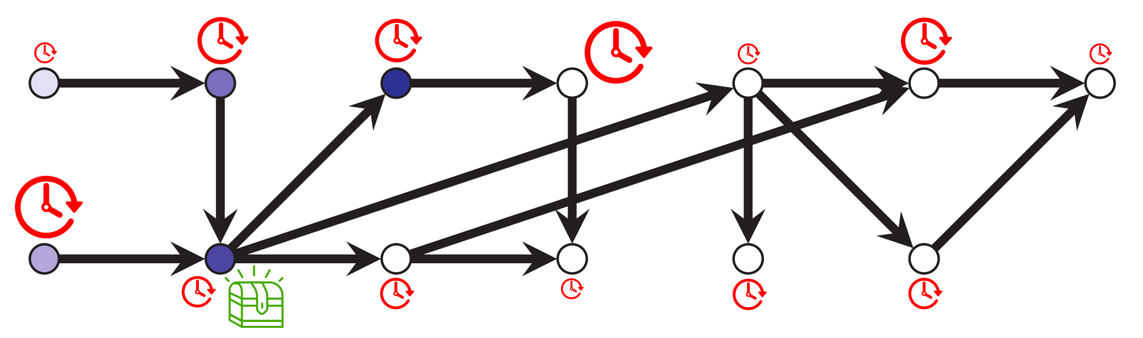
Setting: At each round, **hider** randomly located at some vertex of $\mathcal{G}$.

Goal: Search it.

Constraint: Can examine a vertex only if all in-neighbors already examined (precedence constraints).

Remark: Can **stop** current **search** and go to next round, even if hider was not found.

Instance example:



: precedence constraints
: selected search s
: performed search $\mathbf{s}[\mathbf{W}]$
: random hider $\mathbf{W} \in \{\mathbf{e}_i,\ i \in [n]\}$, $\mathbf{w}^\star \triangleq \mathbb{E}[\mathbf{W}]$
: random cost $\mathbf{C} \in [0,1]^n$, $\mathbf{c}^\star \triangleq \mathbb{E}[\mathbf{C}]$

At each round t, $(\mathbf{W}_t, \mathbf{C}_t) \overset{iid}{\sim} \mathbb{P}_\mathbf{W} \otimes \mathbb{P}_\mathbf{C}$.

Semi-bandit feedback: $W_{i,t}$ and $C_{i,t}$ revealed for each examined vertex i.

Purpose: Design policy π maximizing the expected number of hiders found within **budget** B

$$F_B(\pi) \triangleq \mathbb{E}\left[\sum_{t=1}^{\tau_B - 1} \sum_{i \in \mathbf{s}_t[\mathbf{W}_t]} W_{i,t}\right],$$

τ_B = random round at which remaining budget becomes negative.

Evaluation: **Expected regret**

$$R_B(\pi) \triangleq F_B^\star - F_B(\pi).$$

Contributions

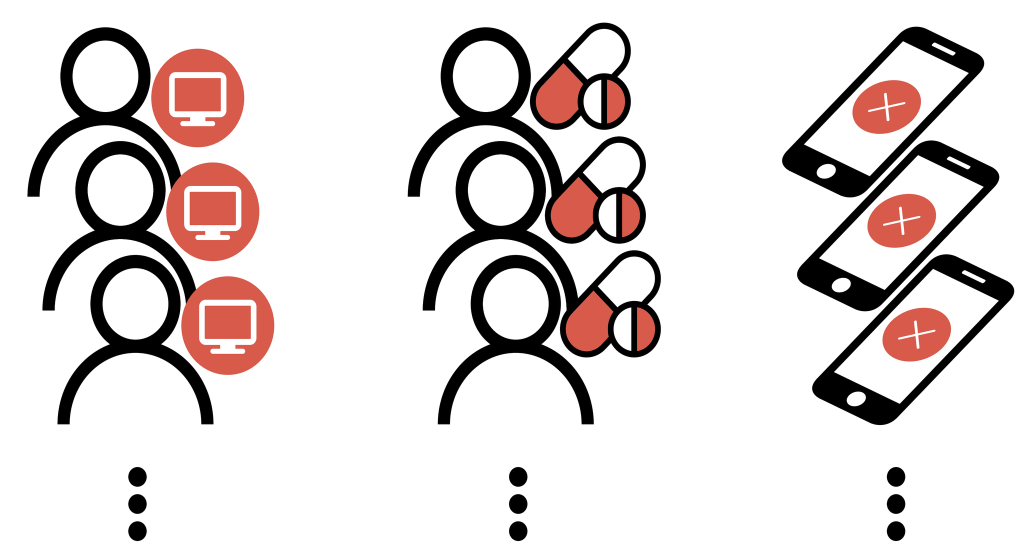
New budgeted bandit setting: Non linear, order dependent.

Offline oracle design: Quasi-optimal, efficient, online-adapted.

Online setting: Variance-based algorithm, upper/lower bounds.

Motivations

Online advertising | **Diagnostics** | **Troubleshooting**



some sequence of actions that generates a conversion from the user.

some medical test revealing the pathology in the patient.

some malfunctioning component in the device [3].

References

[1] Sidney, J. B. (1975). Decomposition Algorithms for Single-Machine Sequencing with Precedence Relations and Deferral Costs. *Operations Research*, 23(2):283–298.

[2] Xia, Y., Qin, T., Ma, W., Yu, N., and Liu, T.-Y. (2016b). Budgeted multi-armed bandits with multiple plays. In *International Joint Conference on Artificial Intelligence*.

[3] Jensen, F. V., Kjaerulff, U., Kristiansen, B., Langseth, H., Skaanning, C., Vomlel, J., and Vomlelova, M. (2001). The SACSO methodology for troubleshooting complex systems. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, 15(4):321–333.

[4] Wang, Q. and Chen, W. (2017). Improving regret bounds for combinatorial semi-bandits with probabilistically triggered arms and its applications. In *Neural Information Processing Systems*.

[5] Audibert, J. Y., Munos, R., and Szepesvári, C. (2009). Exploration-exploitation tradeoff using variance estimates in multi-armed bandits. *Theoretical Computer Science*, 410(19):1876–1902.

Offline search-and-stop

Vectors $\mathbf{w}^\star, \mathbf{c}^\star$ are given as input.

$$J(\mathbf{s}) \triangleq \frac{\sum_{i=1}^{|\mathbf{s}|} c^\star_{s_i}\left(1 - \sum_{j=1}^{i-1} w^\star_{s_j}\right)}{\sum_{i=1}^{|\mathbf{s}|} w^\star_{s_i}}, \quad J(\mathbf{s}^\star) = J^\star \triangleq \min_{\mathbf{s}} J(\mathbf{s}).$$

$J(\mathbf{s})$ = **ratio** between **expected cost paid** and **expected number of hiders found**, on a single round, selecting search s.

Proposition (based on [2]). *Stationary strategy $\pi^\star = (\mathbf{s}^\star, \mathbf{s}^\star, \dots)$ is* ***quasi-optimal****:* $\frac{B-n}{J^\star} \le F_B(\pi^\star) \le F_B^\star \le \frac{B+n}{J^\star}$.

Consequence:

$$R_B(\pi) \simeq \sum_{t=1}^{T_B} \Delta(\mathbf{s}_t), \quad \text{where}$$

$$\Delta(\mathbf{s}) \triangleq \tfrac{1}{J^\star} \sum_{i=1}^{|\mathbf{s}|} c^\star_{s_i}\left(1 - \sum_{j=1}^{i-1} w^\star_{s_j}\right) - \sum_{i=1}^{|\mathbf{s}|} w^\star_{s_i} \ge 0.$$

$$T_B \triangleq 2B / c_{\min}.$$

$c_{\min} \triangleq$ lower bound on expected cost payed for a round.

Remark: $J(\mathbf{s})$ also = expected cost paid to find a **single hider**, following strategy $(\mathbf{s}, \mathbf{s}, \dots)$.

How to minimize J?

Fixed support

How to minimize J over orderings?

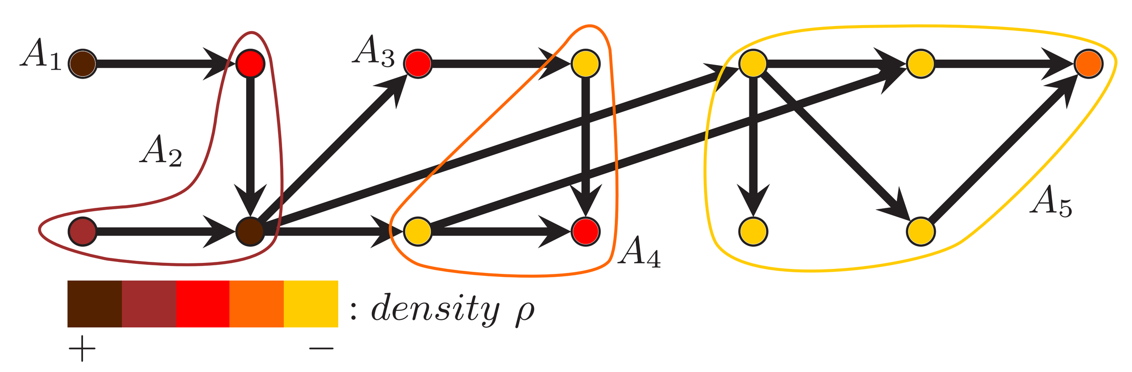
Definition (density). $\rho(A) \triangleq \frac{\sum_{i \in A} w^\star_i}{\sum_{i \in A} c^\star_i}$.

Proposition (order property). *If $\rho(\mathbf{a}) \ge \rho(\mathbf{b})$, then*

$$J(\mathbf{ab}) \le J(\mathbf{ba}).$$

Definition (Sidney's decomposition [1]). $A_1 \sqcup A_2 \sqcup \dots \sqcup A_k = [n]$ *s.t.* $\forall i \in [k]$, A_i = *maximum density search in* $\mathcal{G}\langle A_i \sqcup \dots \sqcup A_k \rangle$.

Example:



Theorem. *The optimal ordering is consistent with the Sidney's decomposition.*

Fixed order

Assume the optimal ordering s is known.

What is the optimal search $\mathbf{s}^\star$?

Proposition (support property). *If* $\rho(\mathbf{z}) \ge \rho(\mathbf{y})$, *then*

$$J(\mathbf{xy}) \ge \min\{J(\mathbf{x}), J(\mathbf{xyz})\}$$

Theorem. *The minimizer of J is of the form* $\mathbf{s}^\star = (s_1, \dots, s_i)$.

Generalization to $\sum w_i \neq 1$

Assume only estimates $(\mathbf{w}, \mathbf{c}) \in \mathbb{R}^2_+$ of $\mathbf{w}^\star, \mathbf{c}^\star$ are available.

Idea: Replace J by J^+.

Advantage:

$$J(\mathbf{s}; \mathbf{w}, \mathbf{c})^+ \le J(\mathbf{s}; \mathbf{w}^\star, \mathbf{c}^\star) = J(\mathbf{s}),$$

if $\mathbf{w} \ge \mathbf{w}^\star$ and $\mathbf{c} \le \mathbf{c}^\star$.

Theorem. *The minimizer of J^+ is of the form* $\mathbf{s}^\star = (s_1, \dots, s_i)$.

Online search-and-stop

Vectors $\mathbf{w}^\star, \mathbf{c}^\star$ are unknown.

$$N_{\mathbf{w},i,t-1} \triangleq \sum_{u=1}^{t-1} \mathbb{I}\{i \in \mathbf{s}_u\}, \quad \bar{w}_{i,t-1} \triangleq \frac{\sum_{u=1}^{t-1} \mathbb{I}\{i \in \mathbf{s}_u\} W_{i,u}}{N_{\mathbf{w},i,t-1}},$$

$$N_{\mathbf{c},i,t-1} \triangleq \sum_{u=1}^{t-1} \mathbb{I}\{i \in \mathbf{s}_u[\mathbf{W}_u]\}, \quad \bar{c}_{i,t-1} \triangleq \frac{\sum_{u=1}^{t-1} \mathbb{I}\{i \in \mathbf{s}_u[\mathbf{W}_u]\} C_{i,u}}{N_{\mathbf{c},i,t-1}}.$$

$$c_{i,t} \triangleq \left(\bar{c}_{i,t-1} - \sqrt{\frac{0.5\zeta \log t}{N_{\mathbf{c},i,t-1}}}\right)^+,$$

$$w_{i,t} \triangleq \min\left\{\bar{w}_{i,t-1} + \sqrt{\frac{2\zeta \bar{w}_{i,t-1}(1 - \bar{w}_{i,t-1}) \log t}{N_{\mathbf{w},i,t-1}}} + \frac{3\zeta \log t}{N_{\mathbf{w},i,t-1}}, 1\right\}.$$

Algorithm CUCB-V for sequential search-and-stop

Input: $\mathcal{G}$.
for $t = 1..\infty$ **do**
select $\mathbf{s}_t = \textsc{Oracle}(\mathbf{w}_t, \mathbf{c}_t, \mathcal{G}) = \arg\min J^+(\,\cdot\,; \mathbf{w}_t, \mathbf{c}_t)$.
perform $\mathbf{s}_t[\mathbf{W}_t]$.
collect feedback: update counters and empirical averages.
end for

$$\Delta_{i,\min} \triangleq \inf_{\mathbf{s} \neq \mathbf{s}^\star:\ i \in \mathbf{s}} \Delta(\mathbf{s}) > 0.$$

Theorem (upper bound).

$R_B(\pi_{\text{CUCB-V}}) =$

$$\mathcal{O}\left(n \log T_B \sum_{i \in [n]} \frac{1 + (J^\star + n)^2 \sigma_i^2}{{J^\star}^2 \Delta_{i,\min}} + \frac{(J^\star + n)}{J^\star} \log\left(\frac{n}{J^\star \Delta_{i,\min}}\right)\right).$$

In addition,

$$\sup R_B(\pi) = \mathcal{O}\left(\sqrt{n}\left(1 + \frac{n}{J^\star}\right)\sqrt{T_B \log T_B}\right),$$

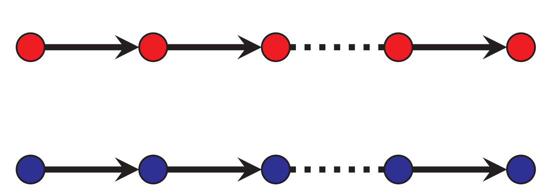
where sup *over all possible sequential search-and-stop problems with fixed $c_{\min}$ and $J^\star$.*

Theorem (lower bound). *On some sequential search-and-stop problem, the optimal online policy π satisfies*

$$-4 + \frac{1}{28}\sqrt{\frac{B}{n}} \le R_B(\pi) = \mathcal{O}\left(\sqrt{B \log\left(\frac{B}{n}\right)}\right).$$

path **a**

path **b**



Experiments