UniRank: Unimodal Bandit Algorithm for Online Ranking

Camille-Sovanneary Gauthier, **Romaric Gaudel**, Elisa Fromont

romaric.gaudel@irisa.fr

ICML, July 2022

institut
universitaire
de France

(1) Unimodality allows the design of efficient bandit algorithms by focusing on a small set of sub-optimal arms.

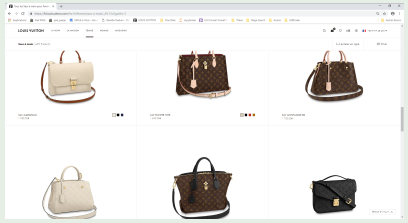
(2) Unimodality is a flexible property.

(1) Unimodality allows the design of efficient bandit algorithms by focusing on a small set of sub-optimal arms.

(2) Unimodality is a flexible property.

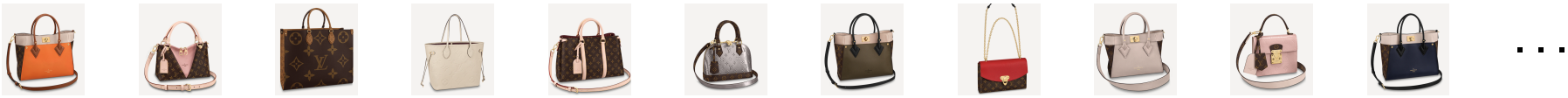
Example on a combinatorial bandit problem

Online Learning to Rank
aka. Multiple-Play Bandit



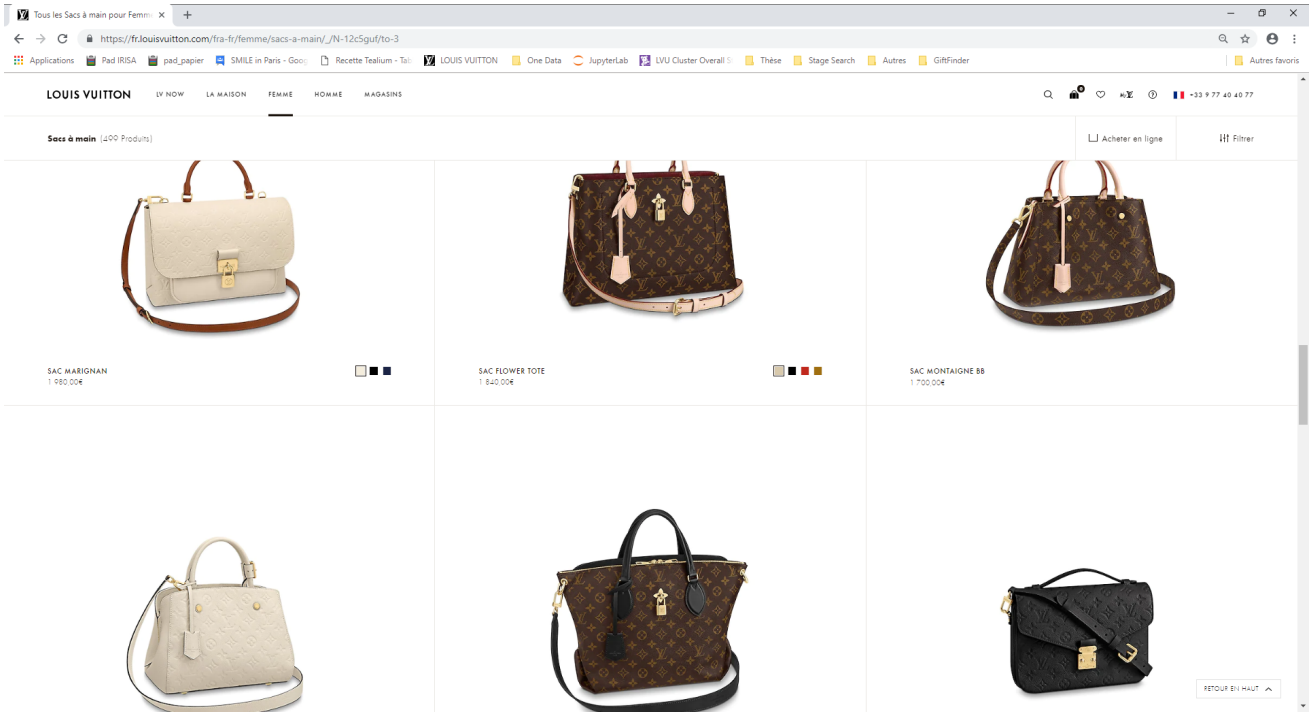
Multiple Recommendations

- Choose K items among $L \gg K$



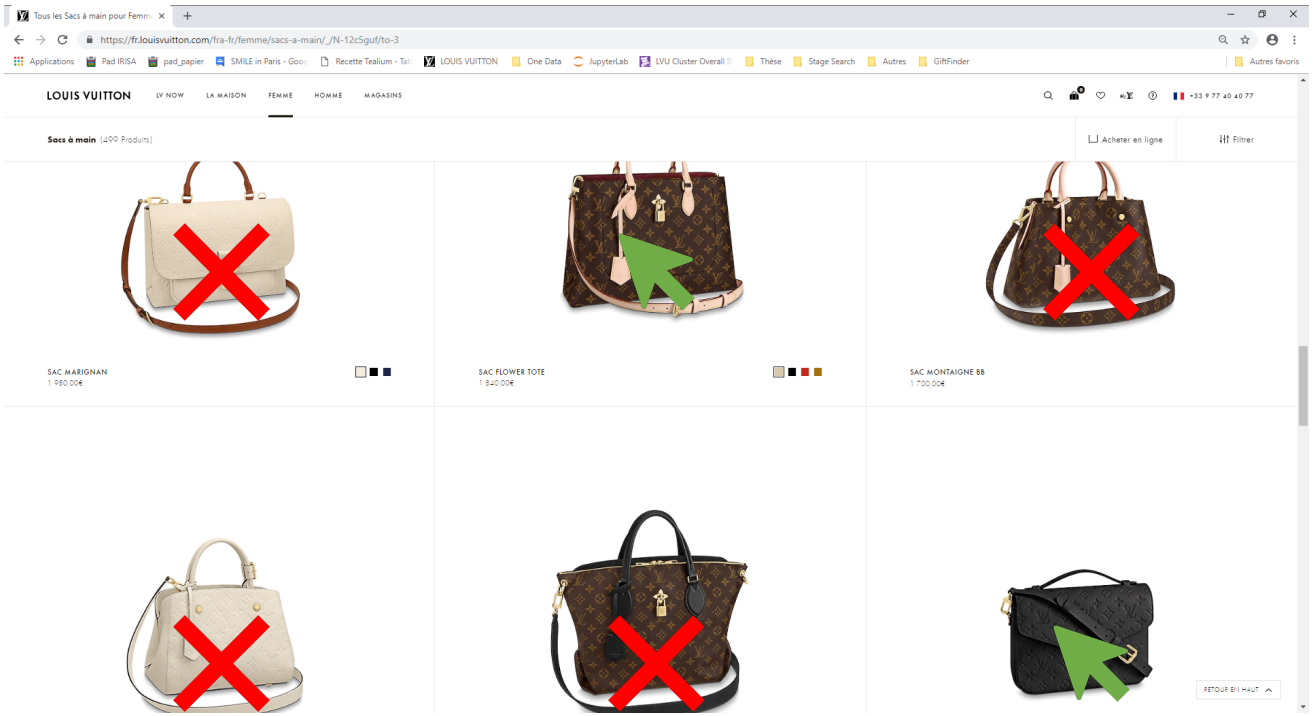
Multiple Recommendations

- Choose K items among $L \gg K$



Multiple Recommendations

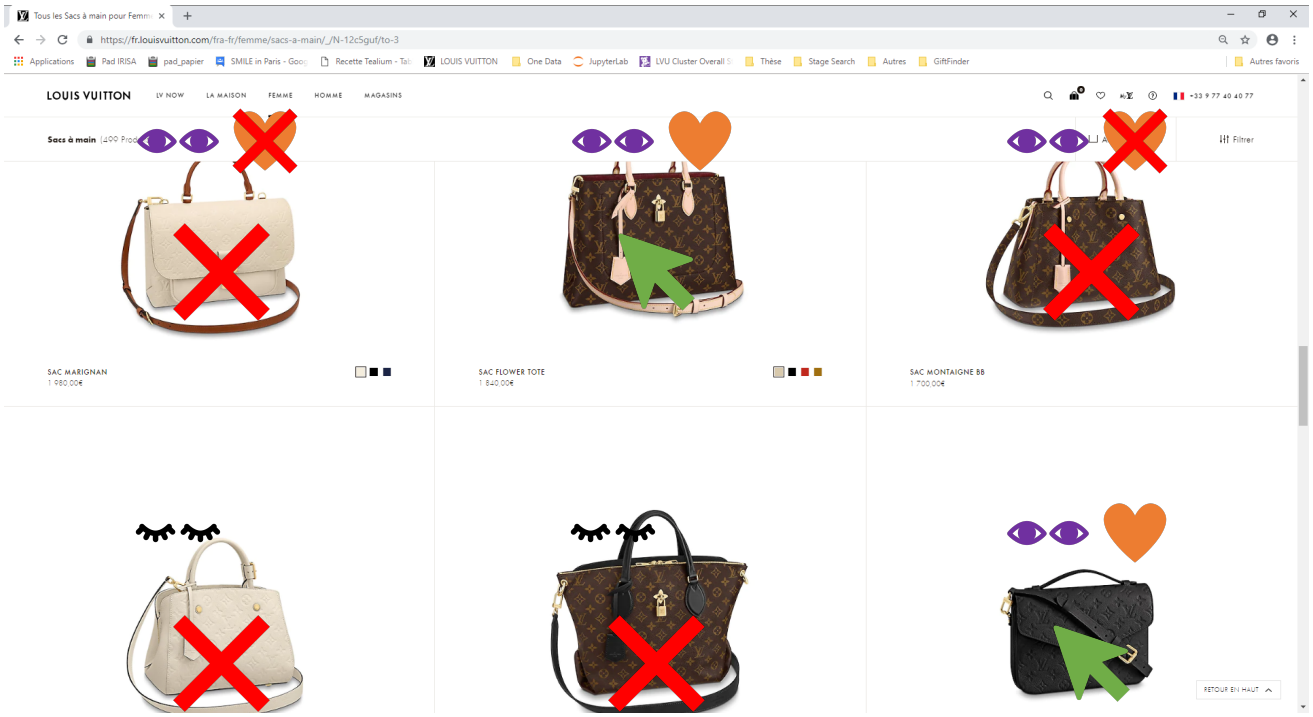
- Choose K items among $L \gg K$



- Get feedback: $(c_1, \ldots, c_K)$
  - to be maximized: $\sum_k c_k$

Multiple Recommendations

- Choose K items among $L \gg K$

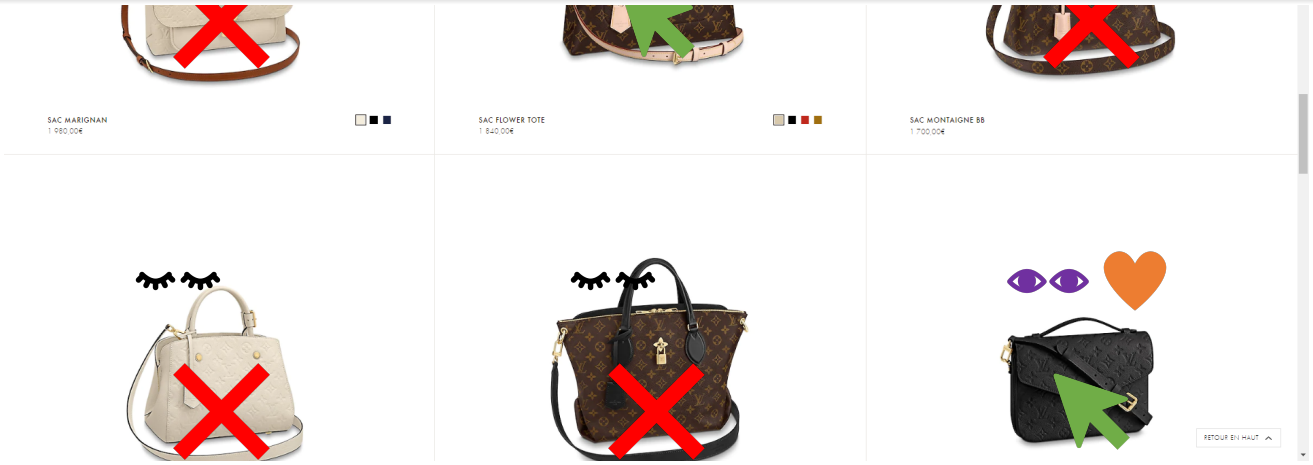


- Get feedback: $(c_1, \ldots, c_K)$
  - to be maximized: $\sum_k c_k$
- Click = look at & like
- Best recommendation: assign most attractive items to most seen positions

Multiple Recommendations

State of the art (Information Retrieval)

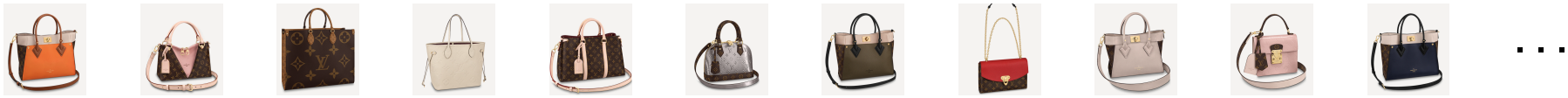
- Model the behavior of users & infer parameters of the model from logs
  - PBM
  - CM
  - DCM
  - ...



- Get feedback: $(c_1, \ldots, c_K)$
  - to be maximized: $\sum_k c_k$
- Click = look at & like
- Best recommendation: assign most attractive items to most seen positions

Online Learning to Rank

- Choose L items among $K \gg L$
  - Identify the L best items (clicked with highest probability)



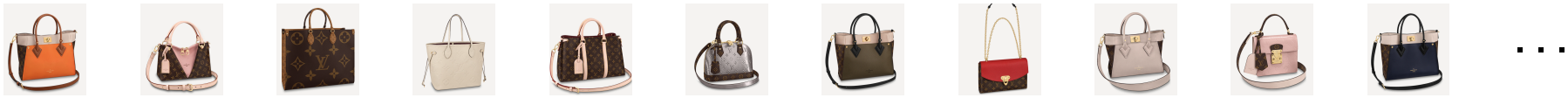
Logs

recommendation $\boldsymbol{a}(t)$

Reco. syst. → User

Online Learning to Rank

- Choose L items among $K \gg L$
  - Identify the L best items (clicked with highest probability)



Logs

recommendation $\boldsymbol{a}(t)$

Reco. syst. → User

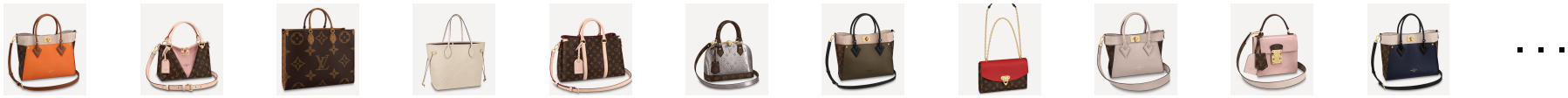
feedback $\boldsymbol{c}(t)$

reward $r(t) = \sum_k c_k(t)$ of expectation $\mu_{\boldsymbol{a}(t)}$

ENSAI

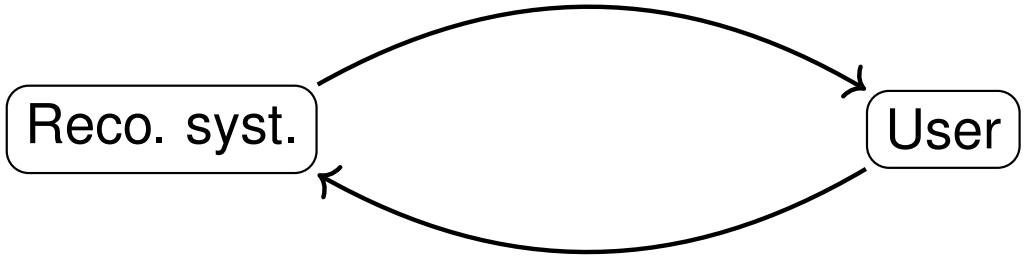
Online Learning to Rank

- Choose L items among $K \gg L$
  - Identify the L best items (clicked with highest probability)



Logs

recommendation $\boldsymbol{a}(t+1)$



Online Learning to Rank

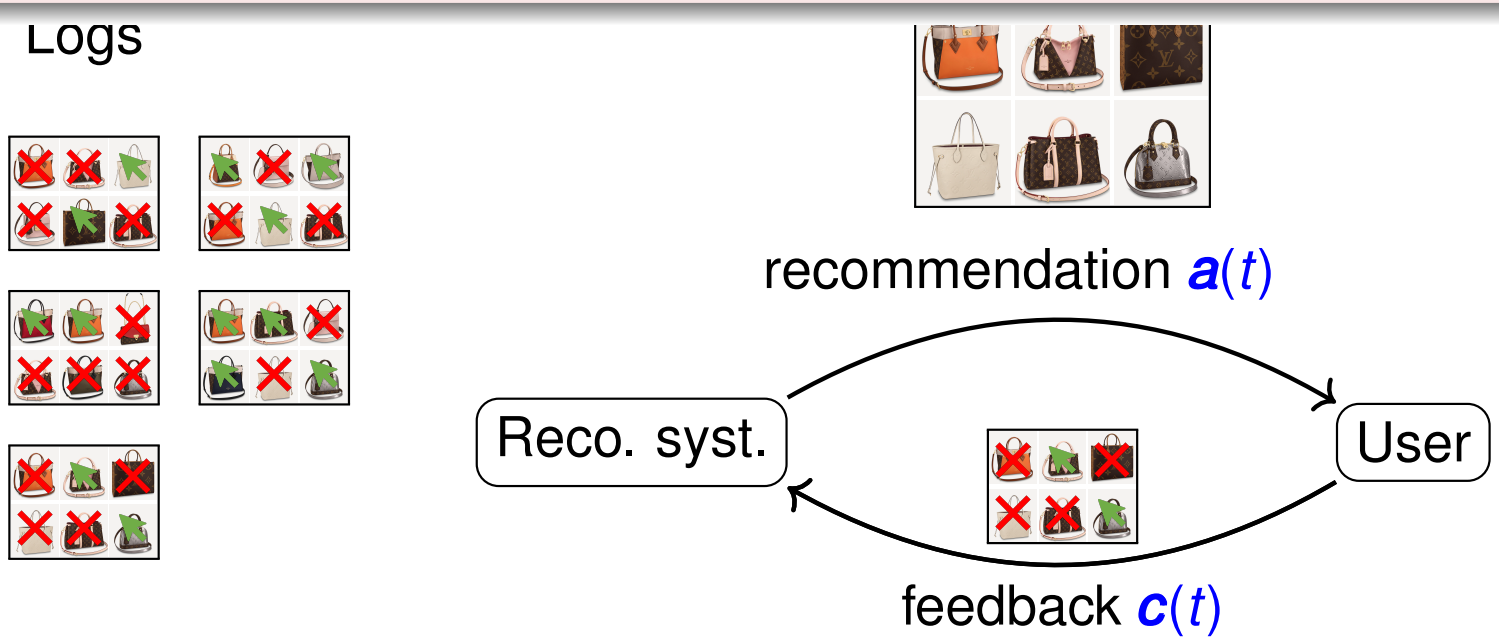
\o/ A Multi-Armed Bandit problem !

- Available data depend on past recommendation

$\Longrightarrow$ balance ***exploration*** and ***exploitation***

- Aim at minimizing the expectation of the ***regret***
  $R(T) = T\mu^* - \sum_{t=1}^{T} \mu_{\boldsymbol{a}(t)}$, where $\mu^* = \max_{\boldsymbol{a}} \mu_{\boldsymbol{a}}$

...



State of the Art for Bandit Setting

ALGORITHM	CLICK MODEL	REGRET
CascadeKL-UCB (Kveton et al., 2015)	CM	$\mathcal{O}\left((L-K)/\Delta \log T\right)$
PBM-PIE (Lagree et al., 2016)	PBM (κ known)	$\mathcal{O}\left((L-K)/\Delta \log T\right)$
PMED-Hinge (Komiyama et al., 2017)	PBM	$\mathcal{O}\left(C^* \log T\right)$
S-GRAB (Us, ICML'21)	PBM*, PBM	$\mathcal{O}\left(KL/\Delta \log T\right)$
GRAB (Us, ICML'21)	PBM*	$\mathcal{O}\left(L/\Delta \log T\right)$
	PBM	$\mathcal{O}\left(KL/\Delta \log T\right)^a$
PB-MHB (Us, IDA'21)	PBM	X
BC-MPTS (Komiyama et al., 2015)	PBM (κ known)	X
PBM-TS (Lagree et al., 2016)	PBM (κ known)	X

[a] conjecture

ENSAI

State of the Art for Bandit Setting

ALGORITHM	CLICK MODEL	REGRET
CascadeKL-UCB (Kveton et al., 2015)	CM	$\mathcal{O}\left((L-K)/\Delta \log T\right)$
PBM-PIE (Lagree et al., 2016)	PBM (κ known)	$\mathcal{O}\left((L-K)/\Delta \log T\right)$
PMED-Hinge (Komiyama et al., 2017)	PBM	$\mathcal{O}\left(C^* \log T\right)$
S-GRAB (Us, ICML'21)	PBM*, PBM	$\mathcal{O}\left(KL/\Delta \log T\right)$
GRAB (Us, ICML'21)	PBM*	$\mathcal{O}\left(L/\Delta \log T\right)$
	PBM	$\mathcal{O}\left(KL/\Delta \log T\right)^a$
PB-MHB (Us, IDA'21)	PBM	X
BC-MPTS (Komiyama et al., 2015)	PBM (κ known)	X
PBM-TS (Lagree et al., 2016)	PBM (κ known)	X

[a] conjecture

ENSAI

State of the Art for Bandit Setting

ALGORITHM	CLICK MODEL	REGRET
CascadeKL-UCB (Kveton et al., 2015)	CM	$\mathcal{O}\left((L-K)/\Delta \log T\right)$
PBM-PIE (Lagree et al., 2016)	PBM (κ known)	$\mathcal{O}\left((L-K)/\Delta \log T\right)$
PMED-Hinge (Komiyama et al., 2017)	PBM	$\mathcal{O}\left(C^* \log T\right)$
S-GRAB (Us, ICML'21)	PBM*, PBM	$\mathcal{O}\left(KL/\Delta \log T\right)$
GRAB (Us, ICML'21)	PBM*	$\mathcal{O}\left(L/\Delta \log T\right)$
	PBM	$\mathcal{O}\left(KL/\Delta \log T\right)^a$
PB-MHB (Us, IDA'21)	PBM	X
BC-MPTS (Komiyama et al., 2015)	PBM (κ known)	X
PBM-TS (Lagree et al., 2016)	PBM (κ known)	X
TopRank (Lattimore et al. 2018)	PBM, CM, ...	$\mathcal{O}\left(KL/\Delta \log T\right)$

[a] conjecture

ENSAI

State of the Art for Bandit Setting

ALGORITHM	CLICK MODEL	REGRET
CascadeKL-UCB (Kveton et al., 2015)	CM	$\mathcal{O}((L-K)/\Delta \log T)$
PBM-PIE (Lagree et al., 2016)	PBM (κ known)	$\mathcal{O}((L-K)/\Delta \log T)$
PMED-Hinge (Komiyama et al., 2017)	PBM	$\mathcal{O}(C^* \log T)$
S-GRAB (Us, ICML'21)	PBM*, PBM	$\mathcal{O}(KL/\Delta \log T)$
GRAB (Us, ICML'21)	PBM*	$\mathcal{O}(L/\Delta \log T)$
	PBM	$\mathcal{O}(KL/\Delta \log T)^a$
PB-MHB (Us, IDA'21)	PBM	X
BC-MPTS (Komiyama et al., 2015)	PBM (κ known)	X
PBM-TS (Lagree et al., 2016)	PBM (κ known)	X
TopRank (Lattimore et al. 2018)	PBM, CM, ...	$\mathcal{O}(KL/\Delta \log T)$
UniRank (Us, ICML'22) (ingredients: Unimodality + focus on partitions on items + Learning to Rank statistic)	CM	$\mathcal{O}((L-K)/\Delta \log T)$
	PBM*, ...	$\mathcal{O}(L/\Delta \log T)$
	PBM, CM, ...	$\mathcal{O}(KL/\Delta \log T)^a$

[a] conjecture

ENSAI

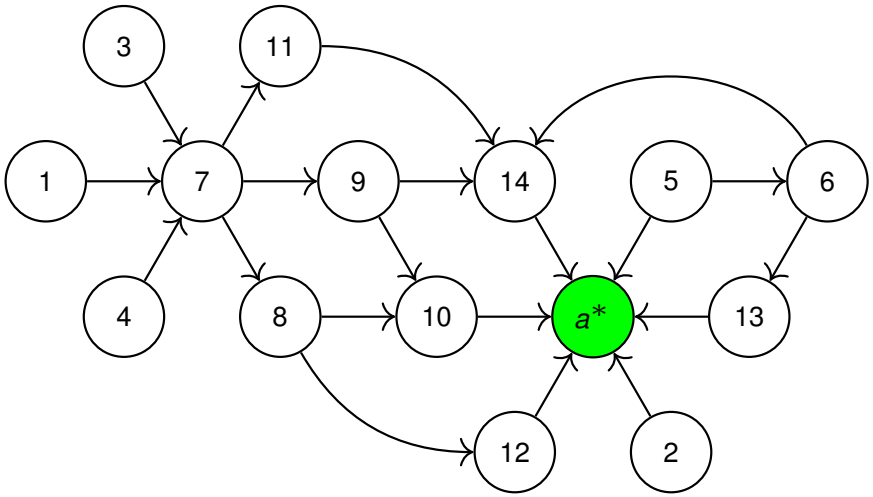
Background: Unimodal Bandit Setting & OSUB

Combes and Proutière. Unimodal bandits: Regret lower bounds and optimal algorithms, ICML'14.

- Hypothesis
  - unique optimal arm: $\text{argmax}_a \mu_a = \{a^*\}$
  - $\forall a \neq a^*, \exists a^+ \in \mathcal{N}(a), \mu_{a^+} > \mu_a$
  - with $\mathcal{N}(a)$ the set of neighbors of a in a known graph G of degree γ

- Typical theoretical result

$$R(T) = \mathcal{O}\left(\sum_{a \in \mathcal{N}(a^*)} \frac{\log T}{\mu^* - \mu_a}\right)$$
$$= \mathcal{O}\left(\frac{\gamma}{\Delta} \log T\right)$$



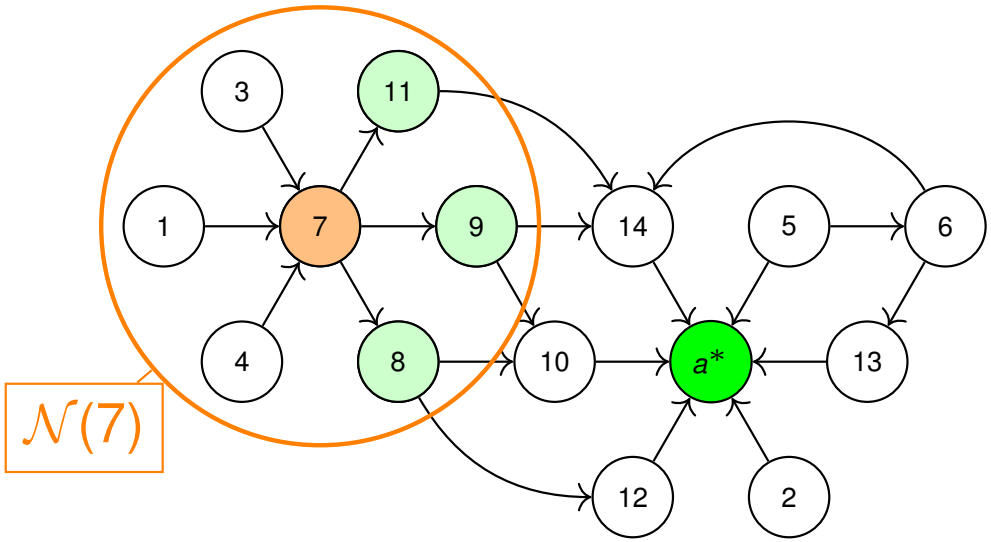
Background: Unimodal Bandit Setting & OSUB

Combes and Proutière. Unimodal bandits: Regret lower bounds and optimal algorithms, ICML'14.

- Hypothesis
  - unique optimal arm: $\text{argmax}_a \mu_a = \{a^*\}$
  - $\forall a \neq a^*, \exists a^+ \in \mathcal{N}(a), \mu_{a^+} > \mu_a$
  - with $\mathcal{N}(a)$ the set of neighbors of a in a known graph G of degree γ

- Typical theoretical result

$$R(T) = \mathcal{O}\left(\sum_{a \in \mathcal{N}(a^*)} \frac{\log T}{\mu^* - \mu_a}\right)$$
$$= \mathcal{O}\left(\frac{\gamma}{\Delta} \log T\right)$$



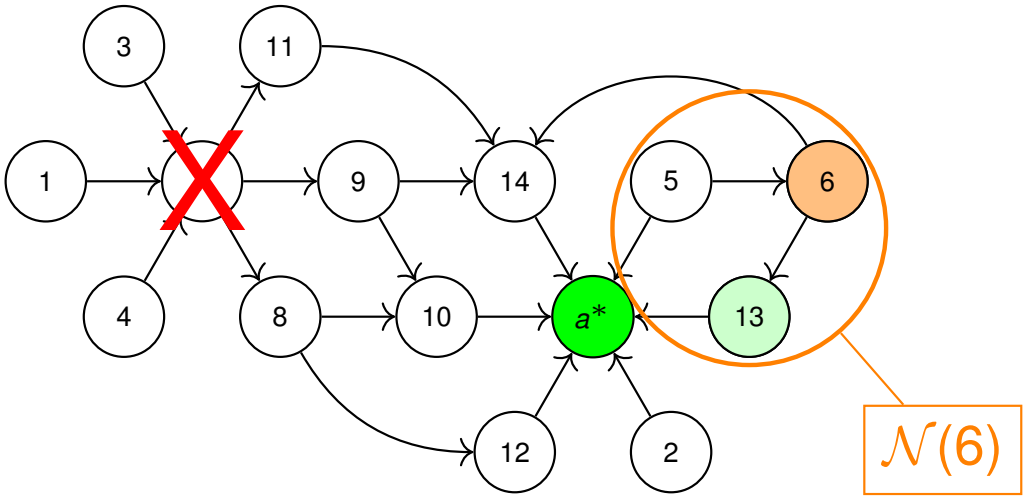
Background: Unimodal Bandit Setting & OSUB

Combes and Proutière. Unimodal bandits: Regret lower bounds and optimal algorithms, ICML'14.

- Hypothesis
  - unique optimal arm: $\text{argmax}_a \mu_a = \{a^*\}$
  - $\forall a \neq a^*, \exists a^+ \in \mathcal{N}(a), \mu_{a^+} > \mu_a$
  - with $\mathcal{N}(a)$ the set of neighbors of a in a known graph G of degree γ

- Typical theoretical result

$$R(T) = \mathcal{O}\left(\sum_{a \in \mathcal{N}(a^*)} \frac{\log T}{\mu^* - \mu_a}\right)$$
$$= \mathcal{O}\left(\frac{\gamma}{\Delta} \log T\right)$$



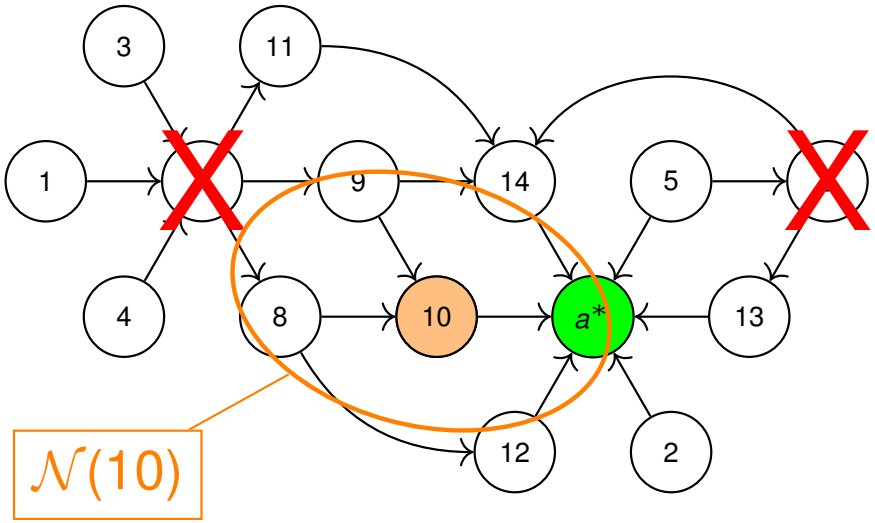
Background: Unimodal Bandit Setting & OSUB

Combes and Proutière. Unimodal bandits: Regret lower bounds and optimal algorithms, ICML'14.

- Hypothesis
  - unique optimal arm: $\text{argmax}_a \mu_a = \{a^*\}$
  - $\forall a \neq a^*, \exists a^+ \in \mathcal{N}(a), \mu_{a^+} > \mu_a$
  - with $\mathcal{N}(a)$ the set of neighbors of a in a known graph G of degree γ

- Typical theoretical result

$$R(T) = \mathcal{O}\left(\sum_{a \in \mathcal{N}(a^*)} \frac{\log T}{\mu^* - \mu_a}\right)$$
$$= \mathcal{O}\left(\frac{\gamma}{\Delta} \log T\right)$$



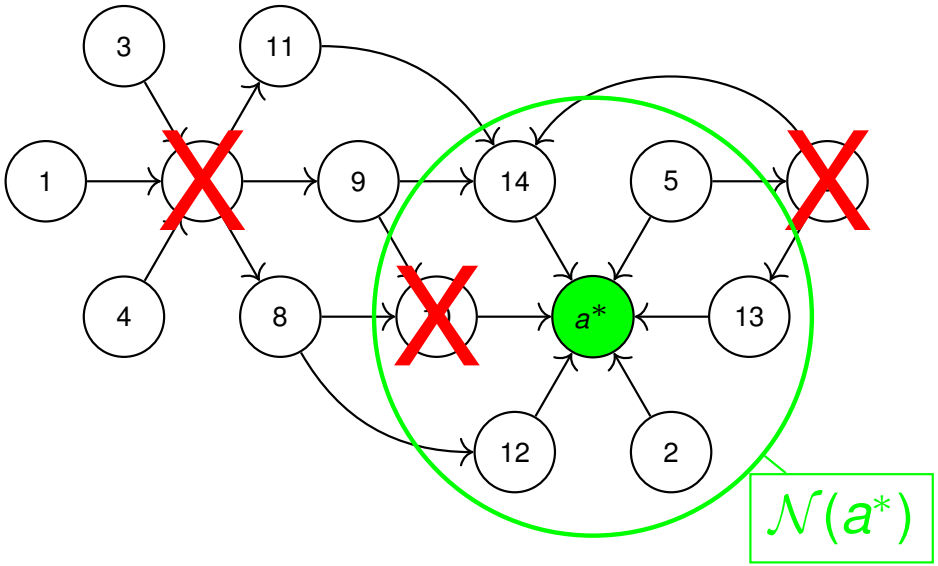
Background: Unimodal Bandit Setting & OSUB

Combes and Proutière. Unimodal bandits: Regret lower bounds and optimal algorithms, ICML'14.

- Hypothesis
  - unique optimal arm: $\text{argmax}_a \mu_a = \{a^*\}$
  - $\forall a \neq a^*, \exists a^+ \in \mathcal{N}(a), \mu_{a^+} > \mu_a$
  - with $\mathcal{N}(a)$ the set of neighbors of a in a known graph G of degree γ

- Typical theoretical result

$$R(T) = \mathcal{O}\left(\sum_{a \in \mathcal{N}(a^*)} \frac{\log T}{\mu^* - \mu_a}\right)$$
$$= \mathcal{O}\left(\frac{\gamma}{\Delta} \log T\right)$$



S-GRAB: OSUB Applied to Online Learning to Rank

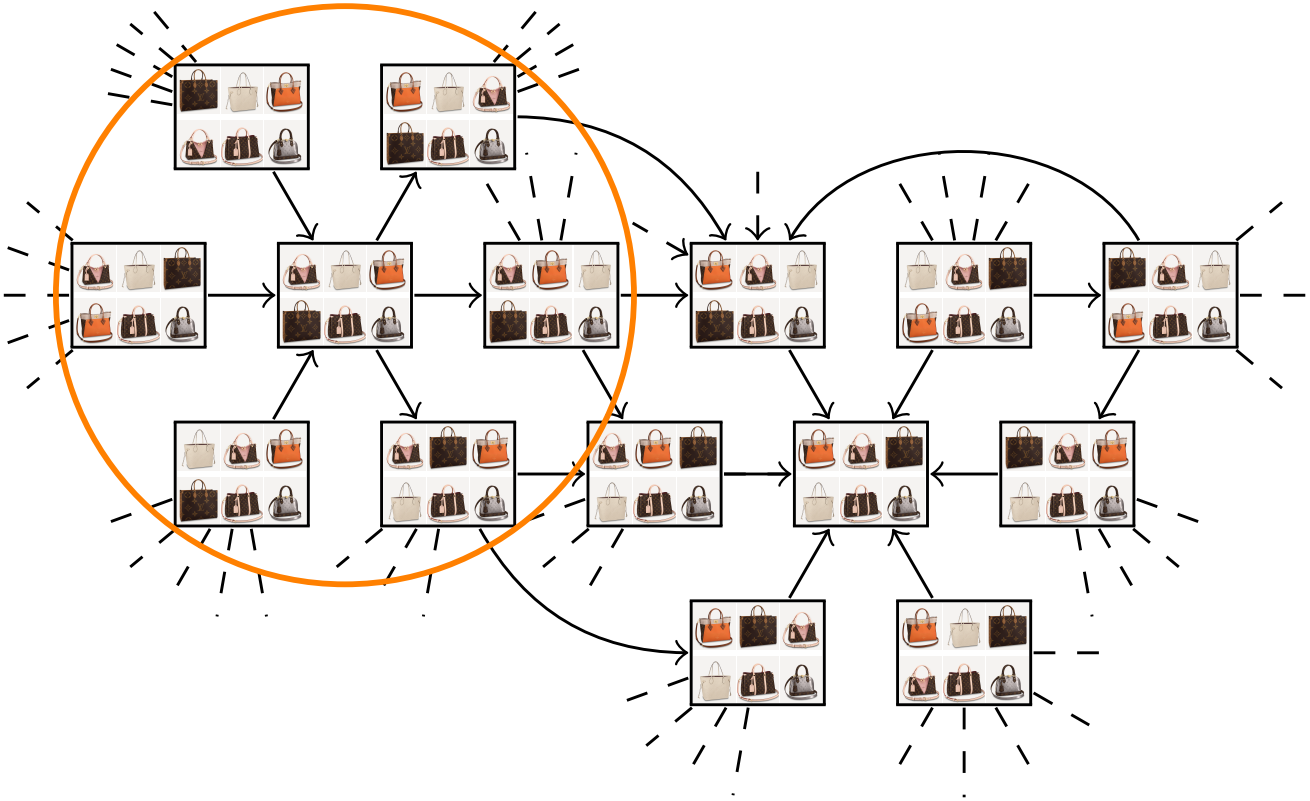
Gauthier, G., Fromont and Lompo. Parametric graph for unimodal ranking bandit, ICML'21.

- Rational: the reward increases when you exchange two items such that the most attractive item gets in the most looked-at position.

$$\mathcal{N}(\boldsymbol{a}) = \left\{\boldsymbol{a} \circ (\ell, \ell') : \ell, \ell' \in [L]^2, \ell > \ell'\right\}$$

$$R(T) = \mathcal{O}\left(\frac{KL}{\Delta} \log T\right)$$

(Same as TopRank)



S-GRAB: OSUB Applied to Online Learning to Rank

GRAB (Gauthier, G., Fromont and Lompo, 2021)

- Parameterized neighborhoods of size L
- Parameter learned on the fly

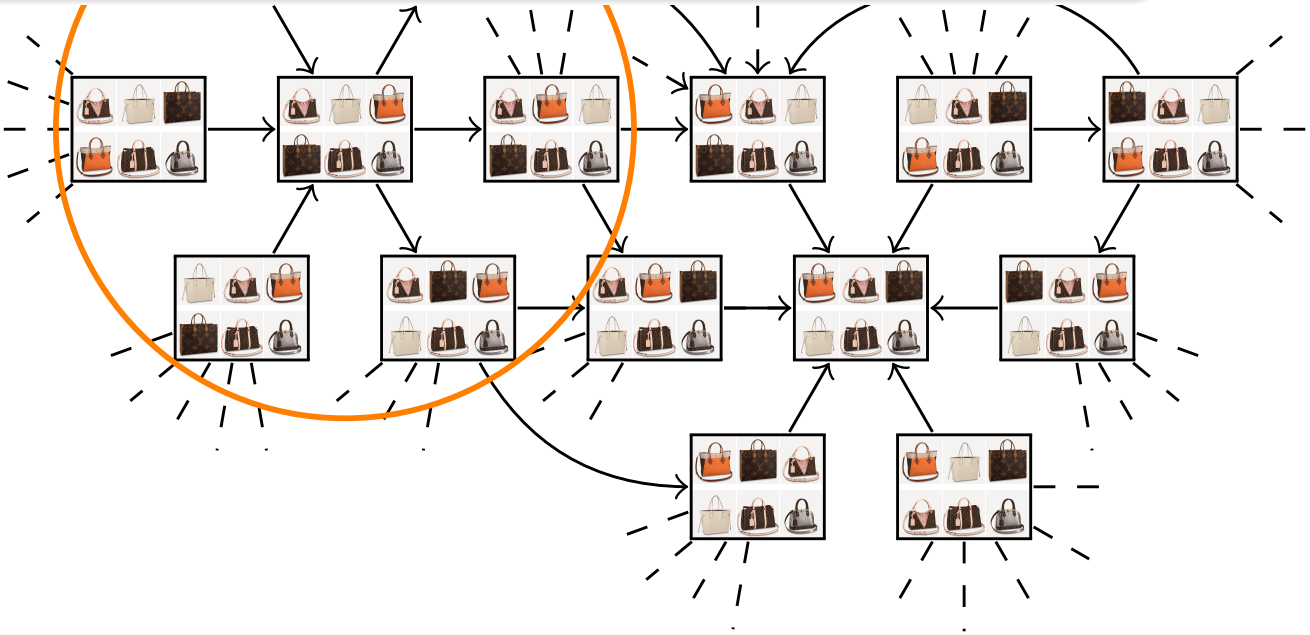
$\Longrightarrow$ $R(T) = \mathcal{O}\left(\frac{L}{\Delta} \log T\right)$

andit, ICML'21.

items such that
on.

$\mathcal{O}\left(\frac{KL}{\Delta} \log T\right)$

as TopRank)



ENSAI

S-GRAB: OSUB Applied to Online Learning to Rank

GRAB (Gauthier, G., Fromont and Lompo, 2021)

- Parameterized neighborhoods of size L
- Parameter learned on the fly

$\Longrightarrow$ $R(T) = \mathcal{O}\left(\frac{L}{\Delta} \log T\right)$

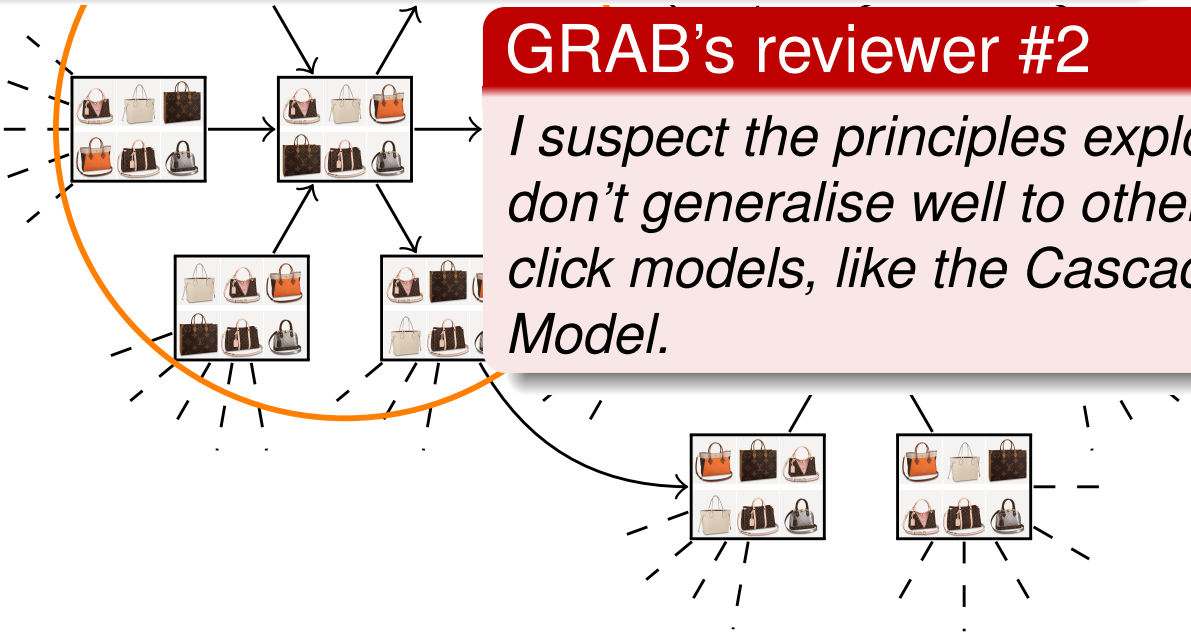
- Based on strong properties of PBM
  - Unicity of the optimal arm
  - Natural parametrization of size KL:
    $\rho_{i,k} = \mathbb{P}(\text{click on item } i \text{ when displayed at position } k)$

andit, ICML'21.

items such that
on.

$\mathcal{O}\left(\frac{KL}{\Delta} \log T\right)$

as TopRank)



GRAB's reviewer #2

I suspect the principles explored here don't generalise well to other popular click models, like the Cascading Click Model.

ENSAI

Multiple-Play as Online Learning to Rank

- A learning to Rank problem
  - Rank the item from the most attractive to the less attractive
  - Attractivity: $\theta_i = \mathbb{P}(c_i \mid \text{the user look at } i)$

Assumptions 3.1* and 3.2 ($\approx$ Learning to Rank)

1. There exists a *strict total order on top-k items* $\succ$
2. Any recommendation $\boldsymbol{a}$ which displays items in an order **coherent** with $\succ$ has **maximal expected reward**

$\Longrightarrow$ unicity (of the best order)

ENSAI

UniRank's Graph

$$\{1,2\} \succ \{3\} \succ \{4,5\} \succ \{6,7\}$$

UniRank's Graph

$$\{1,2\} \succ \{3\} \succ \{4,5\} \succ \{6,7\}$$

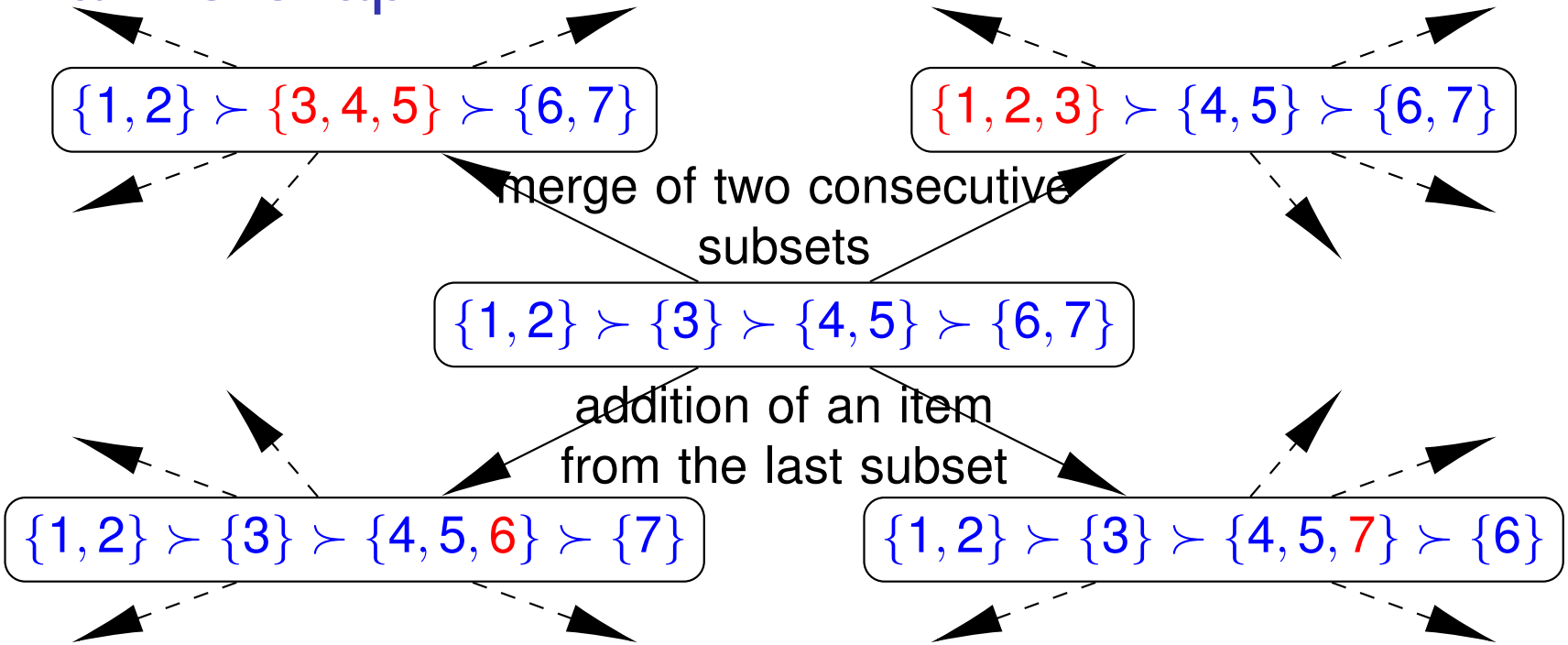
The *ordered partition* $\{1,2\} \succ \{3\} \succ \{4,5\} \succ \{6,7\}$ represents the set of *compatible* recommendations (of size 4)

- $[1,2,3,4]$
- $[1,2,3,5]$
- $[2,1,3,4]$
- $[2,1,3,5]$

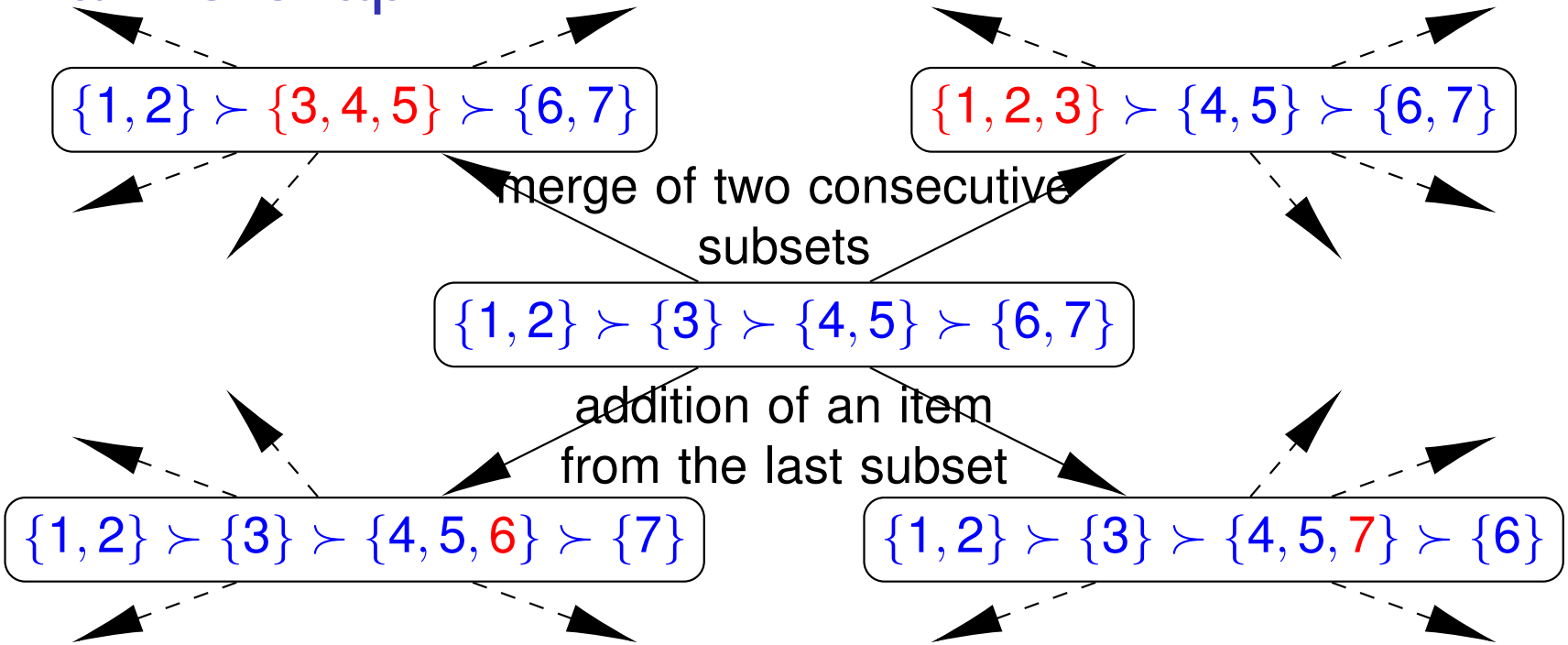
Implies comparison of

- 1 vs. 2
- 4 vs. 5

UniRank's Graph



UniRank's Graph



Lemma 5.4

This neighborhood is sufficient to inspect the "optimality" of $\tilde{P}$

Let (L, K, ρ) be an OLR problem satisfying Assumptions 3.1*, 3.2 and 3.3 and such that $1 \succ 2 \succ \cdots \succ K \succ [L] \setminus [K]$. Denoting $\boldsymbol{P}^* = (\{1\}, \ldots, \{K\}, [L] \setminus [K])$ the optimal partition associated to this order, for any ordered partition of the items

-
-

Relaxed Unimodality

A sub-optimal node may have **no better node** in its neighborhood
(sub-optimality detected by playing itself)

UniRank's Statistics

Expected Click Difference (Lattimore, Kveton, Li, Szepesvári, 2018)

$$\tilde{\Delta}_{i,j}(\boldsymbol{a}) = \mathbb{E}_{\boldsymbol{a}(t)\sim\mathcal{U}(\{\boldsymbol{a},(i,j)\circ\boldsymbol{a}\})}\left[c_i(t) - c_j(t) \mid c_i(t) \neq c_j(t)\right]$$

(strong link with Dueling Bandits)

$$\tilde{\delta}_{i,j}(\boldsymbol{a}) = \mathbb{P}_{\boldsymbol{a}(t)\sim\mathcal{U}(\{\boldsymbol{a},(i,j)\circ\boldsymbol{a}\})}\left[c_i(t) \neq c_j(t)\right]$$

ENSAI

UniRank's Statistics

Expected Click Difference (Lattimore, Kveton, Li, Szepesvári, 2018)

$$\tilde{\Delta}_{i,j}(\boldsymbol{a}) = \mathbb{E}_{\boldsymbol{a}(t)\sim\mathcal{U}(\{\boldsymbol{a},(i,j)\circ\boldsymbol{a}\})}\left[c_i(t) - c_j(t) \mid c_i(t) \neq c_j(t)\right]$$

(strong link with Dueling Bandits)

$$\tilde{\delta}_{i,j}(\boldsymbol{a}) = \mathbb{P}_{\boldsymbol{a}(t)\sim\mathcal{U}(\{\boldsymbol{a},(i,j)\circ\boldsymbol{a}\})}\left[c_i(t) \neq c_j(t)\right]$$

Assumption 3.3

3. For any couple of items (i,j) s.t. $i \succ j$, and for any recommendation $\boldsymbol{a}$ displaying i or j or both,

$$\tilde{\delta}_{i,j}(\boldsymbol{a}) \neq 0 \quad \text{and} \quad \tilde{\Delta}_{i,j}(\boldsymbol{a}) > 0$$

UniRank's Statistics

Lemma 3.1

CM and PBM* models fulfill Assumptions 3.1*, 3.2 and 3.3.

E **Lattimore, Kveton, Li, Szepesvári, 2018)**

$$\tilde{\Delta}_{i,j}(\boldsymbol{a}) = \mathbb{E}_{\boldsymbol{a}(t)\sim\mathcal{U}(\{\boldsymbol{a},(i,j)\circ\boldsymbol{a}\})}\left[c_i(t) - c_j(t) \mid c_i(t) \neq c_j(t)\right]$$

(strong link with Dueling Bandits)

$$\tilde{\delta}_{i,j}(\boldsymbol{a}) = \mathbb{P}_{\boldsymbol{a}(t)\sim\mathcal{U}(\{\boldsymbol{a},(i,j)\circ\boldsymbol{a}\})}\left[c_i(t) \neq c_j(t)\right]$$

Assumption 3.3

3. For any couple of items (i,j) s.t. $i \succ j$, and for any recommendation $\boldsymbol{a}$ displaying i or j or both,

$$\tilde{\delta}_{i,j}(\boldsymbol{a}) \neq 0 \quad \text{and} \quad \tilde{\Delta}_{i,j}(\boldsymbol{a}) > 0$$

ENSAI

UniRank

Unimodal Bandit Algorithm for Online Ranking

Require: number of items L, number of positions K

1: **for** $t = 1, 2, \ldots$ **do**

2: compute the leader partition $\tilde{\boldsymbol{P}}(t)$ given $\hat{S}_{i,j}(t)$ [a]

7: **end for**

---

[a] $O_{i,j}(s) \overset{def}{=} \mathbb{1}\left\{\exists c, (i,j) \in P_c(s)^2\right\} \mathbb{1}\{c_i(s) \neq c_j(s)\}$ $\quad T_{i,j}(t) \overset{def}{=} \sum_{s=1}^{t-1} O_{i,j}(s)$

$\hat{S}_{i,j}(t) \overset{def}{=} \frac{1+\sum_{s=1}^{t-1} O_{i,j}(s)\left(c_i(s)-c_j(s)\right)}{2T_{i,j}(t)}$ related to a mixture of $\left(\frac{1+\tilde{\Delta}_{i,j}(\boldsymbol{a})}{2}\right)_{\boldsymbol{a}\in\mathcal{P}_K^L}$

UniRank

Unimodal Bandit Algorithm for Online Ranking

Require: number of items L, number of positions K

1: **for** $t = 1, 2, \ldots$ **do**

2: compute the leader partition $\tilde{\boldsymbol{P}}(t)$ given $\hat{S}_{i,j}(t)$ [a]

3: $\boldsymbol{P}(t) \leftarrow \tilde{\boldsymbol{P}}(t)$ or $\underset{c \in [\tilde{d}-1]}{\operatorname{argmax}} \; \max_{(i,j) \in \tilde{P}_c(t) \times \tilde{P}_{c+1}(t)} f\left(\hat{S}_{j,i}(t), T_{i,j}(t), \tilde{t}_{\tilde{\boldsymbol{P}}(t)}(t)\right)$ [b]

7: **end for**

---

[a] $O_{i,j}(s) \overset{def}{=} \mathbb{1}\left\{\exists c, (i,j) \in P_c(s)^2\right\} \mathbb{1}\{c_i(s) \neq c_j(s)\}$ $\quad T_{i,j}(t) \overset{def}{=} \sum_{s=1}^{t-1} O_{i,j}(s)$

$\hat{S}_{i,j}(t) \overset{def}{=} \frac{1+\sum_{s=1}^{t-1} O_{i,j}(s)\left(c_i(s)-c_j(s)\right)}{2T_{i,j}(t)}$ $\quad$ related to a mixture of $\left(\frac{1+\tilde{\Delta}_{i,j}(\boldsymbol{a})}{2}\right)_{\boldsymbol{a} \in \mathcal{P}_K^L}$

[b] $f(\hat{\mu}, N, t) \overset{def}{=} \sup\{\mu \in [\hat{\mu}, 1] : N \times \mathrm{kl}(\hat{\mu}, \mu) \leq \log(t) + 3\log(\log(t))\}$

UniRank

Unimodal Bandit Algorithm for Online Ranking

Require: number of items L, number of positions K

1: **for** $t = 1, 2, \ldots$ **do**

2: compute the leader partition $\tilde{\boldsymbol{P}}(t)$ given $\hat{S}_{i,j}(t)$ [a]

3: $\boldsymbol{P}(t) \leftarrow \tilde{\boldsymbol{P}}(t)$ or $\underset{c \in [\tilde{d}-1]}{\operatorname{argmax}} \max_{(i,j) \in \tilde{P}_c(t) \times \tilde{P}_{c+1}(t)} f\left(\hat{S}_{j,i}(t), T_{i,j}(t), \tilde{t}_{\tilde{\boldsymbol{P}}(t)}(t)\right)$ [b]

4: draw the recommendation $\boldsymbol{a}(t)$ uniformly at random in $\mathcal{A}(\boldsymbol{P}(t))$ [c]

5: observe the clicks vector $\boldsymbol{c}(t)$

6: compute $O_{i,j}(t+1)$, $T_{i,j}(t+1)$, and $\hat{S}_{i,j}(t+1)$

7: **end for**

---

[a] $O_{i,j}(s) \overset{def}{=} \mathbb{1}\left\{\exists c, (i,j) \in P_c(s)^2\right\} \mathbb{1}\{c_i(s) \neq c_j(s)\}$ $\quad T_{i,j}(t) \overset{def}{=} \sum_{s=1}^{t-1} O_{i,j}(s)$

$\hat{S}_{i,j}(t) \overset{def}{=} \frac{1+\sum_{s=1}^{t-1} O_{i,j}(s)\left(c_i(s)-c_j(s)\right)}{2T_{i,j}(t)}$ $\quad$ related to a mixture of $\left(\frac{1+\tilde{\Delta}_{i,j}(\boldsymbol{a})}{2}\right)_{\boldsymbol{a} \in \mathcal{P}_K^L}$

[b] $f(\hat{\mu}, N, t) \overset{def}{=} \sup\{\mu \in [\hat{\mu}, 1] : N \times \mathrm{kl}(\hat{\mu}, \mu) \leq \log(t) + 3\log(\log(t))\}$

[c] Recommendations $\boldsymbol{a}$ coherent with $\boldsymbol{P}(t)$

Theoretical Results

Theorem 5.1

Let (L, K, ρ) be an OLR problem satisfying Assumptions 3.1*, 3.2 and 3.3 and such that $1 \succ 2 \succ \cdots \succ K \succ [L] \setminus [K]$. Denoting $\boldsymbol{P}^* = (\{1\}, \ldots, \{K\}, [L] \setminus [K])$ the optimal partition associated to this order, when facing this problem, UniRank fulfills

$$\forall k \in [L]\setminus\{1\},\ \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{1}\left\{ \begin{array}{c} \tilde{\boldsymbol{P}}(t)=\boldsymbol{P}^*, \\ \exists c, P_c(t)=\{\min(k-1,K),k\} \end{array} \right\}\right] \leqslant \frac{16}{\tilde{\delta}_k^* \tilde{\Delta}_k^2} \log T + \mathcal{O}(\log\log T)$$

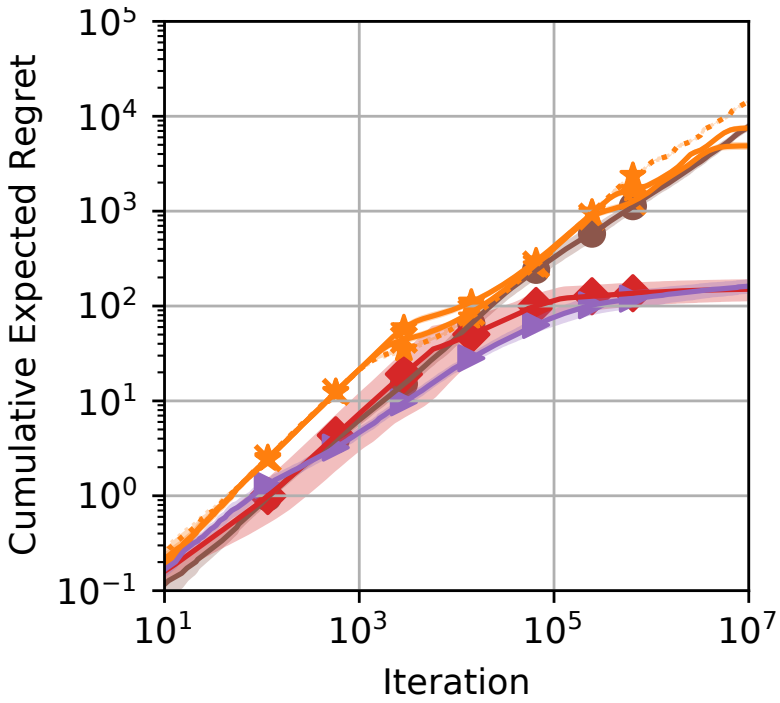
$$\text{and } \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{1}\{\tilde{\boldsymbol{P}}(t) \neq \boldsymbol{P}^*\}\right] = \mathcal{O}(\log\log T),$$

and hence

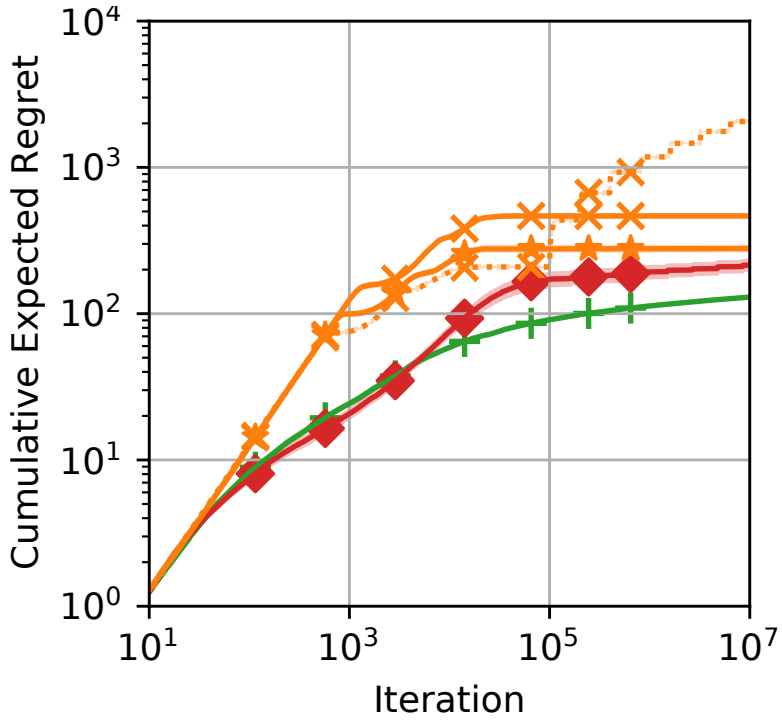
$$R(T) \leqslant \sum_{k=2}^{L} 8 \frac{\Delta_k}{\tilde{\delta}_k^* \tilde{\Delta}_k^2} \log T + \mathcal{O}(\log\log T) = \mathcal{O}\left(\frac{L}{\Delta} \log T\right).$$

Empirical Results

the lower,
the better



Yandex 8107157 (PBM)



Well-chosen θ (CM)

Alg. Dedicated to PBM

- GRAB
- PB-MHB, c=10^3, m=1

Generic Algorithms

- UniRank
- TopRank,T=10^7
- TopRank, T=10^{12}
- TopRank, doubling

Alg. Dedicated to CM

- CascadeKL-UCB

ENSAI

Conclusion

- UniRank: efficient algorithm for online ranking
  - $\mathcal{O}(L/\Delta \log T)$ regret for CM and PBM*
  - Based on unimodality
    - ⋆ Flexible
    - ⋆ Focus on the intrinsic complexity of the problem
  - Based on Learning to Rank point of view
    - ⋆ Generic formulation
  - With a dueling-bandit-like criterion
- Future work
  - Account for contextual information

ENSAI

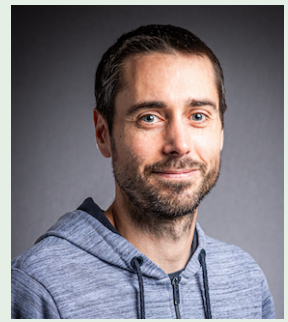
Questions? Remarks? (Multiple) Recommendations?

Poster

Hall E #633

@ ICML'22

Camille-S. G.



Romaric G.

Related work @ Complex feedback in online learning

On Saturday, Room 314 - 315

Theoretical Results for State of the Art Click-Models

Corollary 5.2

With the click-model CM with probability θ_i to click on item i when it is observed, UniRank fulfills

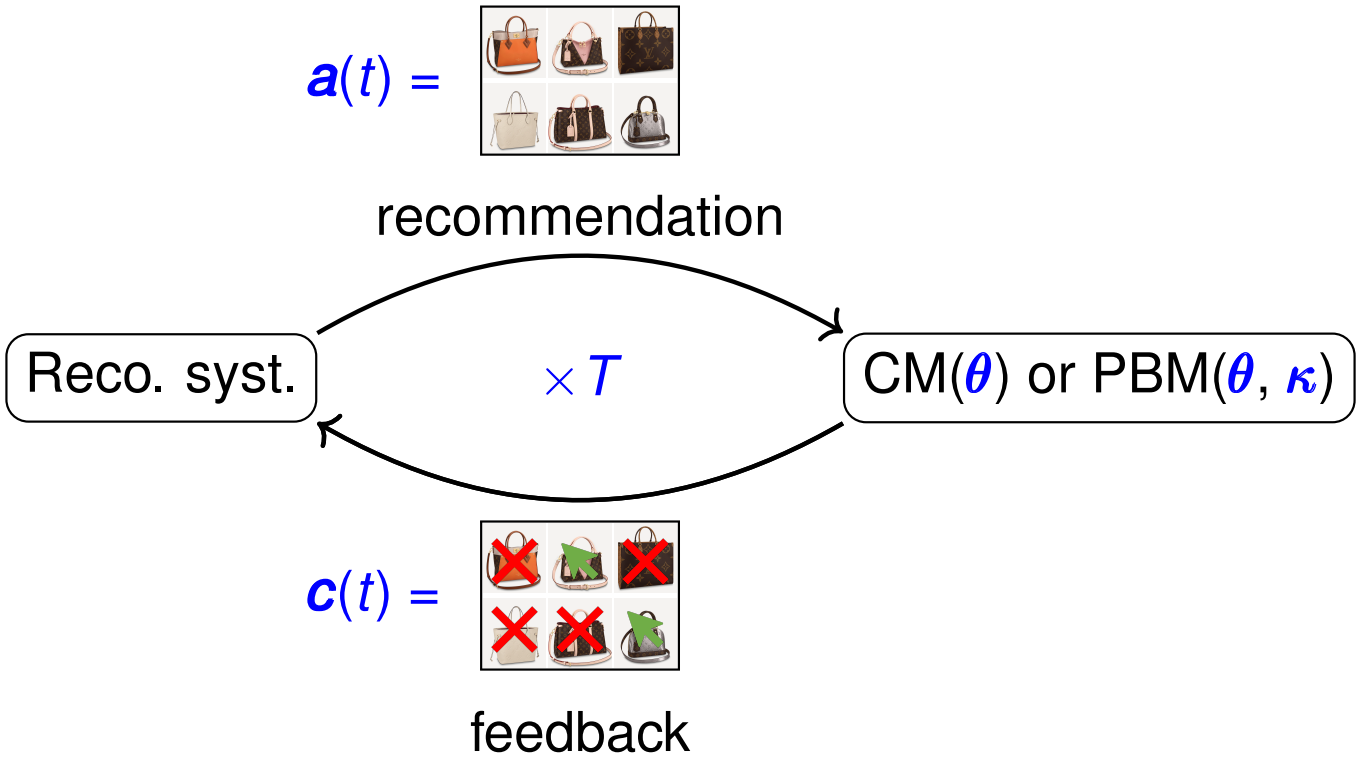
$$R(T) = \mathcal{O}\left((L-K)\frac{\theta_K + \theta_{K+1}}{\theta_K - \theta_{K+1}} \log T\right).$$

Corollary 5.3

With the click-model PBM with probability θ_i to click on item i when it is observed and the probability κ_k of observing the position k, UniRank fulfills

$$R(T) = \mathcal{O}\left(\frac{L}{\Delta} \log T\right).$$

Experimental Setting



- Parameters
  - θ and κ from real-life data or well-chosen
  - $T = 10^7$ iterations
- Score
  - $R_T = T.\mu^* - \sum_{t=1}^{T} \mu_{a(t)}$
  - Averaged on 20 repetitions per model parameters
  - The lower the better

ENSAI