Top-m identification for linear bandits

Clémence Réda, Émilie Kaufmann,
Andrée Delahaye-Duriez

Motivated by drug repurposing, we propose a generic family of algorithms to tackle the identification of the $m \geq 1$ arms with largest means in a linear bandit model.

Fixed-confidence linear Top-*m* identification

The goal is to identify the $m < K$ arms with the largest expected rewards, using fixed time-independent arm contexts. In a linear bandit model, the expected reward μ_a of each arm a is a linear function of context $x_a \in R^N$

$$\text{for any arm } a,\ \mu_a = \theta^T x_a \text{ for some vector } \theta \in R^N$$

A (ε, *m*, δ)-PAC (Probably Approximately Correct) algorithm for fixed-confidence linear Top-*m* returns a subset of *m* arms $\hat{S}^t_m$ at round *t* *s.t.* if $\mu_{(m)}$ is the m^{th} largest mean reward:

$$\text{Prob}(\hat{S}^t_m \subseteq S^{*\varepsilon}) \geq 1-\delta$$

where $S^{*\varepsilon} = \{ a : \mu_a \geq \mu_{(m)} - \varepsilon \}$

Contributions

GIFA (Gap-Index Focused Algorithm) family for Top-*m*

We propose a generic family of algorithms aimed at Top-*m* identification, which encompasses known algorithms for Top-*m* and best arm identification (LUCB[1], UGapE[2], LinGapE[3], ...). These algorithms rely on good gap indices $B_{ij}(t)$ for any arm pair (i, j) *s.t.*

$$\text{Prob}\{ \forall t > 0, \forall j \in (S^{*\varepsilon})^c, \forall k \in S^{*0}, B_{kj}(t) \geq \mu_k - \mu_j \} \geq 1-\delta$$

```
t ← 1
while stopping(t) > ε do
    // estimated m best arms at t
    J(t) ← compute_Jt(t)
    // estimated m-best arm at t
    b_t ← compute_bt(t)
    // challenger to b_t at t
    c_t ← arg max_{a∉J(t)} B_{a,b_t}(t)
    // sampling arms
    I_t ← selection(b_t, c_t)
    r_t ← sample(I_t)
    Update gap indices B_{i,j}(t+1)
    t ← t+1
end while
Ŝ^t_m ← J(t)
return Ŝ^t_m
```

Alg 1: General structure of a GIFA algorithm

* Correctness for partially specified GIFA algorithms:

Theorem 1. If using good gap indices, **(i)** any GIFA algorithm with $b_t = \text{argmax}_{j \in J(t)} \max_{i \in J(t)} B_{ij}(t)$ and the LUCB stopping rule, **or** **(ii)** any GIFA algorithm using $b_t \in J(t)$ with the UGapE stopping rule, is (ε, *m*, δ)-PAC.

* Existence of good gap indices:

Lemma 3. Gap indices using confidence bounds for linear bandits in [4] are good gap indices.

Proposals for linear Top-*m* identification: *m*-LinGapE and LinGIFA

$\mu_i(t)$ is the empirical mean for arm i at time *t*.

Algorithm	compute_Jt	compute_bt	selection	stopping at t
m-LinGapE	$\underset{j \in [K]}{\overset{[m]}{\text{argmax}}}\ \mu_j(t)$	$\underset{j \in J(t)}{\text{argmax}} \underset{i \in (J(t))^c}{\max} B_{ij}(t)$	greedy, optimized	$B_{c_t b_t}(t) \leq \varepsilon$ (**LUCB rule**)
LinGIFA	$\underset{j \in [K]}{\overset{[m]}{\text{argmin}}} \underset{i \neq j}{\overset{m}{\max}} B_{ij}(t)$	$\underset{j \in J(t)}{\text{argmax}} \underset{i \neq j}{\overset{m}{\max}} B_{ij}(t)$	largest variance, greedy	$\underset{j \in J(t)}{\max} \underset{i \neq j}{\overset{m}{\max}} B_{ij}(t) \leq \varepsilon$ (**UGapE rule**)

Tab. 1: New proposals for linear Top-*m* identication

where

Greedy[3] $\text{argmin}_{a \in [K]} (x_{c_t} - x_{b_t})^T (V_{t-1} + x_a x_a^T)^{-1} (x_{c_t} - x_{b_t})$

Optimized[3] $\text{argmax}_{a \in [K],\ w^*a(b_t,c_t) > 0}\ n_a(t)\ ||w^*(b_t,c_t)||_1 / |w^*_a(b_t,c_t)|$

and $w^*(b_t,c_t) = \text{argmin} \{ ||w||_1 \mid w \in R^K : x_{b_t} - x_{c_t} = \Sigma_{a \in [K]} w_a x_a \}$

Largest variance[2] $\text{argmax}_{a \in \{c_t, b_t\}} x_a^T V_t X_a$

and $V_t = \lambda I_N + \Sigma_{a \in [K]} n_a(t) x_a x_a^T$, $n_a(t)$: # of times *a* pulled until round *t*

Unified sample complexity analysis for LUCB-*like* GIFA algorithms

This subclass of algorithms comprises LUCB, LinGapE and *m*-LinGapE (same rules for estimating J(*t*), b_t and stopping). Analysis leads to upper bounds on stopping time of the form

$$C(H) = \inf_{u>0} \{ u > 1 + H \cdot D(\delta,u) + O(K)^* \}$$

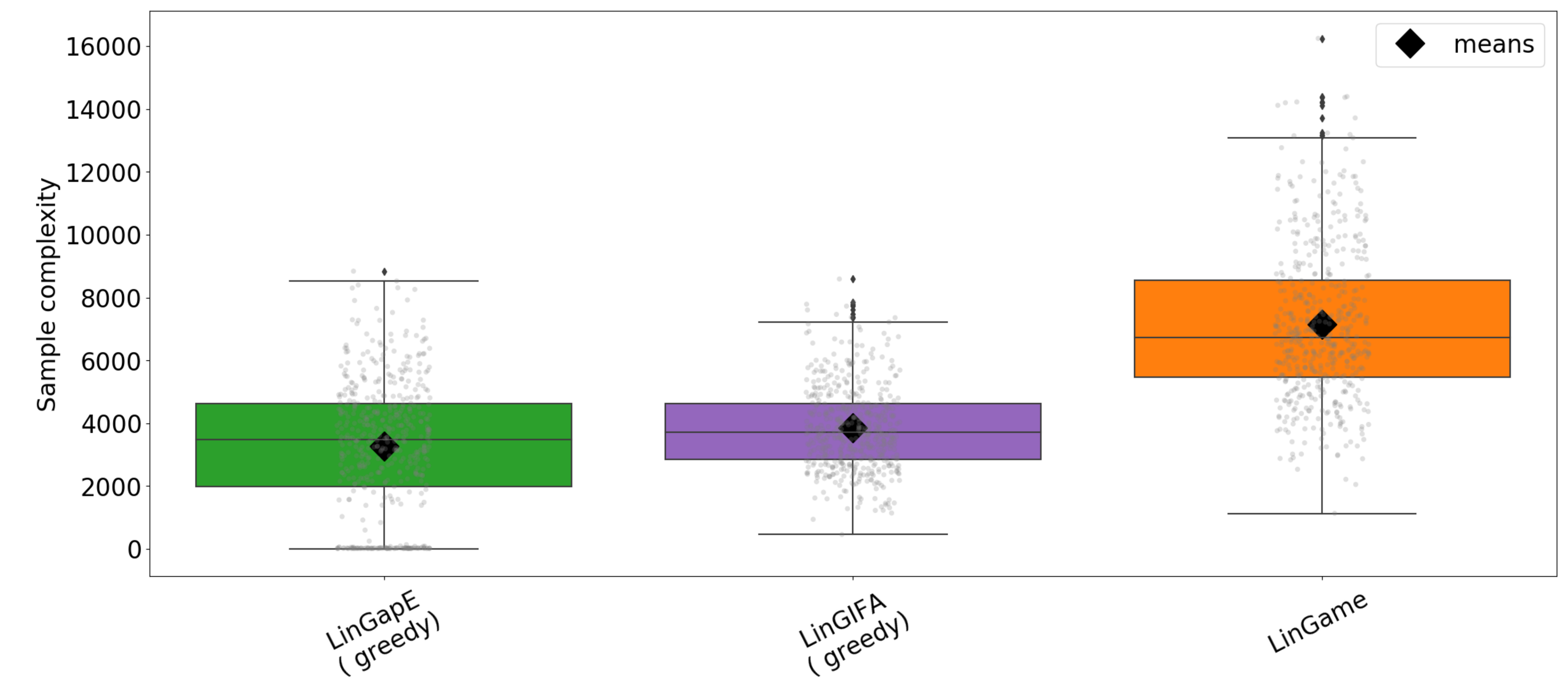
* this term corresponds to an initialization phase where all arms are sampled once, D(δ,u) controls the width of the gap indices

Algorithm	Complexity constant $H^\varepsilon(\mu)$
LUCB	$2\ \Sigma_{a \in [K]} \max(\varepsilon/2, \Delta_a)^{-2}$ $\Delta_a = \mu_a - u_{(m+1)}$ if $a \in S^{*0}$, $\mu_{(m)} - \mu_a$ otherwise
m-LinGapE (largest variance rule)	$4\sigma^2 \Sigma_{a \in [K]} \max(\varepsilon, (\varepsilon + \Delta_a)/3)^{-2}$ σ is the variance on the noise model
m-LinGapE (optimized rule)	$\sigma^2 \Sigma_{a \in [K]} \max_{i,j} \lvert w^*_a(i,j) \rvert (\varepsilon + \max(\Delta_i, \Delta_j)/3)^{-2}$

Tab. 2: Comparaison between upper bounds on stopping time for LUCB-*like* algorithms

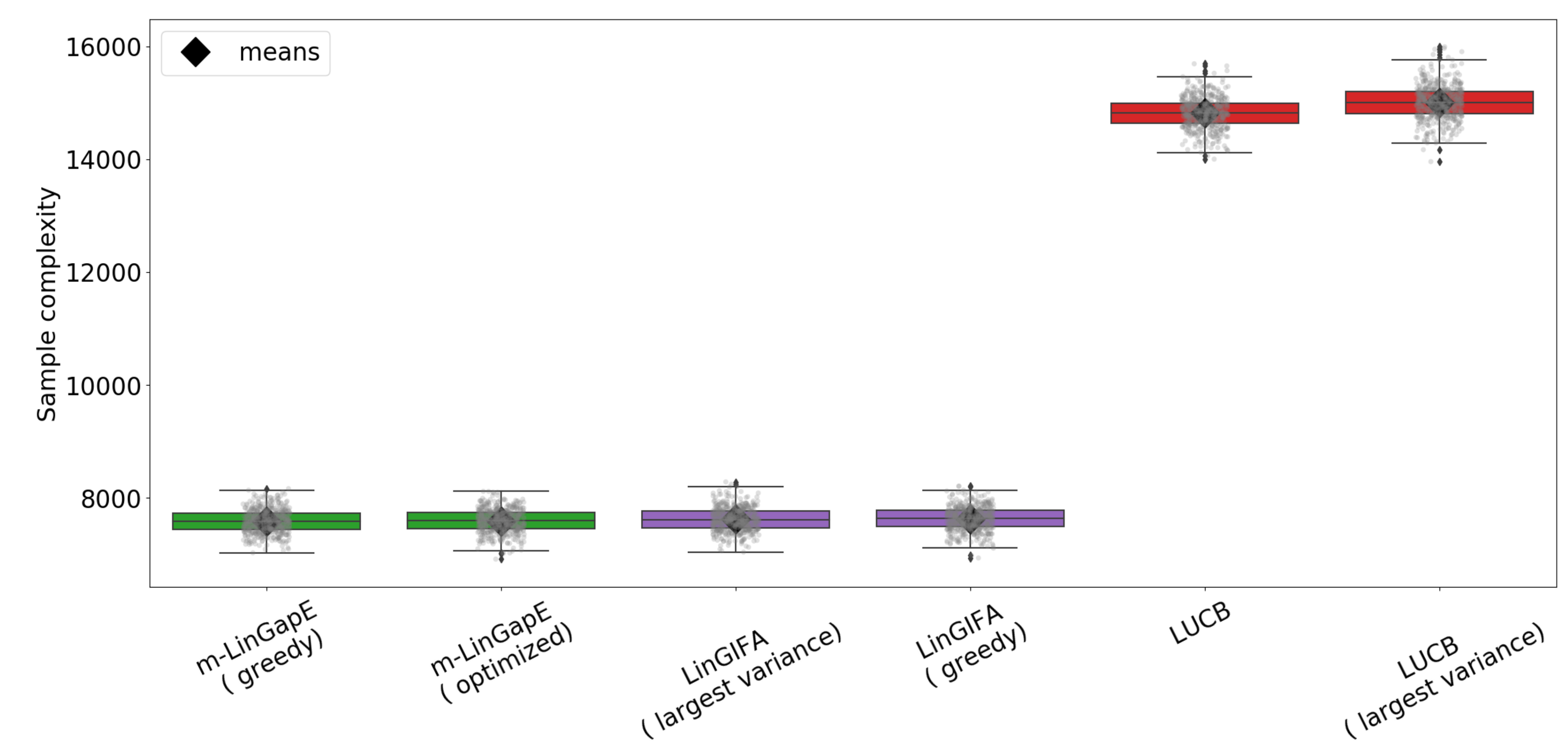
Experimental study (across 500 simulations)

Simulated data: hard best arm (Top-1) identification instance[5]



m-LinGapE and LinGIFA remain competitive even against asymptotically optimal (i.e., when δ ➞ 0) such as LinGame.
(K=3, $\mu_{(1)} - \mu_{(2)} = cos(0.1)$, N=3)

Biological data: application to drug repurposing with *m* > *1*



In a drug repurposing instance for epilepsy (*m* = *5*), sample complexity is improved by a factor ½ compared to LUCB.
(K=10, $\mu_{(m)} - \mu_{(m+1)}$ ~ 0.066, N=71)

References

[1] Kalyanakrishnan, S., Tewari, A., Auer, P., & Stone, P. **"Pac subset selection in stochastic multi-armed bandits"** (2012). ICML.

[2] Gabillon, V., Ghavamzadeh, M., & Lazaric, A. **"Best arm identification: A unified approach to fixed budget and fixed confidence"** (2012). NIPS.

[3] Xu, L., Honda, J., & Sugiyama, M. **"A fully adaptive algorithm for pure exploration in linear bandits"** (2018). AISTATS.

[4] Abbasi-Yadkori Y., Szepesvári C. & Pal D. **"Improved algorithms for linear stochastic bandits"** (2011) NIPS.

[5] Soare M., Lazaric A., & Munos R. **"Best-arm identification in linear bandits"** (2014). NIPS.

[6] Degenne, R., Ménard, P., Shang, X., & Valko, M. **"Gamification of pure exploration for linear bandits"** (2020). ICML.