Best Model Identification: A Rested Bandit Formulation

ICML
International Conference
On Machine Learning

Leonardo Cella[1] Massimiliano Pontil[1,2] Claudio Gentile[3]

[1] CSML, Italian Institute of Technology, Genoa, Italy
[2] Dept. of Computer Science, Univ. College London, UK
[3] Google Research, New York, USA

(Stationary) Best Arm Identification

Stochastic Bandits

A learning policy π sequentially picks one of K options (arms).
Pulled arm yields loss randomly drawn according to an unknown but fixed distribution.

(Stationary) Best Arm Identification

Stochastic Bandits

A learning policy π sequentially picks one of K options (arms).
Pulled arm yields loss randomly drawn according to an unknown but fixed distribution.

BAI Objective: Identify the best arm, the one with smallest expected loss.

Finding the Best Learner

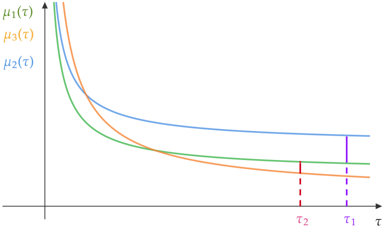
Learners are not static, they tend to improve their skills with experience. Hence, their expected losses are a function of the number of times they have been selected.

Best Model Identification: a Rested-bandit Formulation

- Pulling arm $i \in \mathcal{K} = \{1, \dots, k\}$ at time t, when it was played $\tau = \tau(i, T)$ times, yields random loss with **expectation**:

$$\mu_i(\tau) = \frac{\alpha_i}{\tau^\rho} + \beta_i$$

where $\rho \in (0, 1]$ and $\alpha_i, \beta_i \in \mathbb{R}_{0+}$.

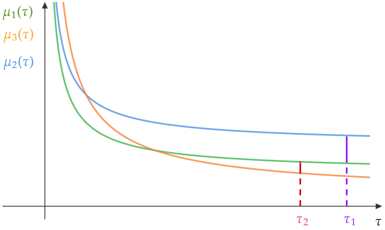


Best Model Identification: a Rested-bandit Formulation

- Pulling arm $i \in \mathcal{K} = \{1, \ldots, k\}$ at time t, when it was played $\tau = \tau(i, T)$ times, yields random loss with **expectation**:

$$\mu_i(\tau) = \frac{\alpha_i}{\tau^\rho} + \beta_i$$

where $\rho \in (0, 1]$ and $\alpha_i, \beta_i \in \mathbb{R}_{0+}$.



- After T interactions π has to commit to one arm $i_{\text{out}} \in \mathcal{K}$.We let $\tau_{\text{out}} = \tau(i_{\text{out}}, T)$ be the number of pulls of i_{out} after T rounds.
- Objective minimize the *pseudo-regret*:

$$R_T(\pi) = \mu_{i_{\text{out}}}(\tau_{\text{out}}) - \mu_{i_T^*}(T)$$

where $i_T^* = \arg\min_{i \in \mathcal{K}} \mu_i(T)$ (notice that $i_{\text{out}}, \tau_{\text{out}}$ are both random variables).

Theoretical Guarantees

Paper outcome:

- We propose an efficient arm-elimination policy;

Theoretical Guarantees

Paper outcome:

- We propose an efficient arm-elimination policy;
- A new trade-off emerges: exploration vs best-arm identification;

Theoretical Guarantees

Paper outcome:

- We propose an efficient arm-elimination policy;
- A new trade-off emerges: exploration vs best-arm identification;
- We prove upper bound on the regret it incurs;

Theoretical Guarantees

Paper outcome:

- We propose an efficient arm-elimination policy;
- A new trade-off emerges: exploration vs best-arm identification;
- We prove upper bound on the regret it incurs;
- We prove a matching lower-bound (up to logarithmic factors) for the case where $K = 2$.

Theoretical Guarantees

Paper outcome:

- We propose an efficient arm-elimination policy;
- A new trade-off emerges: exploration vs best-arm identification;
- We prove upper bound on the regret it incurs;
- We prove a matching lower-bound (up to logarithmic factors) for the case where $K = 2$.
- Our bounds strongly depend on the interplay among parameters $(\alpha_i, \beta_i)_{i=1}^{K}, T$

Theoretical Guarantees

Paper outcome:

- We propose an efficient arm-elimination policy;
- A new trade-off emerges: exploration vs best-arm identification;
- We prove upper bound on the regret it incurs;
- We prove a matching lower-bound (up to logarithmic factors) for the case where $K = 2$.
- Our bounds strongly depend on the interplay among parameters $(\alpha_i, \beta_i)_{i=1}^{K}, T$

Hence, our policy is optimal (up to logs)!


Thank You
For Your Atten