Pareto-efficient Acquisition Functions for Cost-Aware Bayesian Optimization

Gauthier Guinet* †, Valerio Perrone* ‡, Cédric Archambeau ‡

† Massachusetts Institute of Technology ‡ Amazon, Berlin, Germany

gguinet@mit.edu, {vperrone, cedrica}@amazon.com

*joint first author

Bayesian optimization

- **Bayesian optimization** (BO) is a model-based approach to solve the global optimization problem:

$$\min_{\mathbf{x}\in\Omega\subset\mathbb{R}^d} f(\mathbf{x}).$$

- Assumptions: no closed-form expression, no gradient information, and expensive to evaluate.
- BO builds a surrogate model for f, typically a Gaussian process (GP), and loops for a preset number of iterations. Each iteration, it selects a new evaluation point based on an acquisition criterion such as expected improvement (EI).

The cost assumption: Justification, Limits and Heuristics

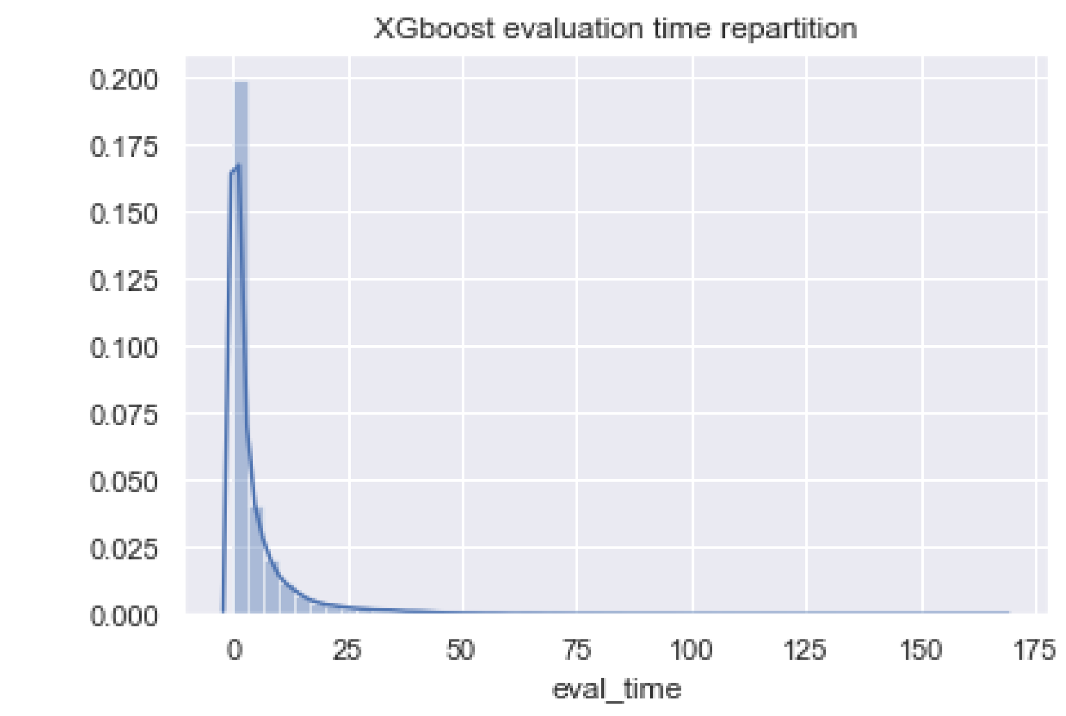


Figure: Density estimation of runtime distribution of 5000 randomly selected points for XGBoost. We witness several orders of magnitude of difference between evaluations.

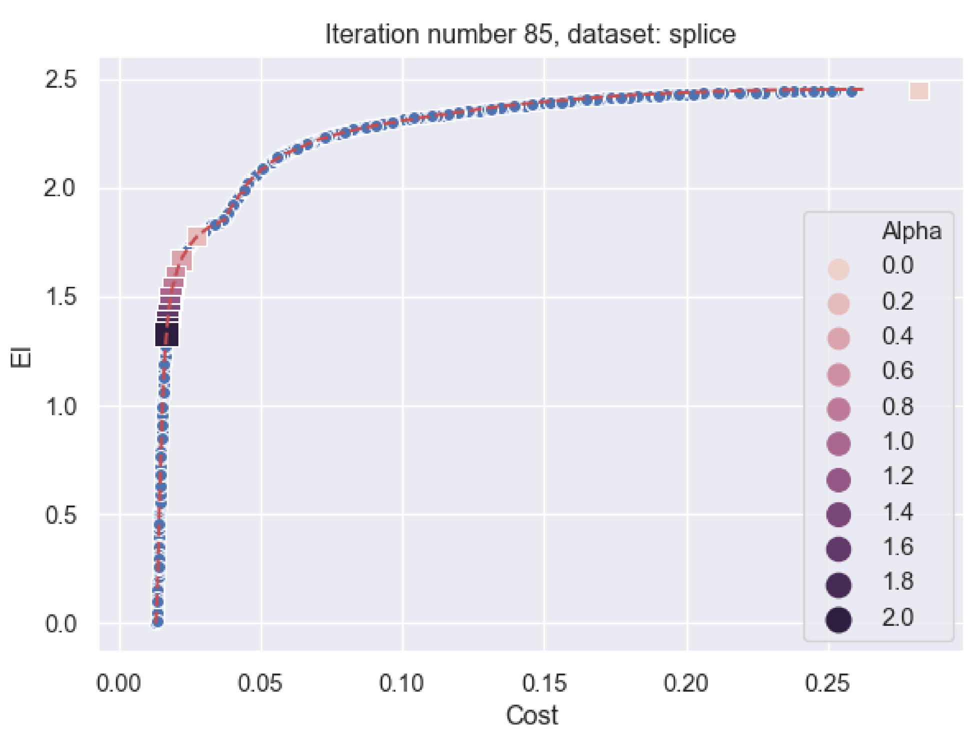
- **The Cost Assumption**: The cost of evaluation f is huge yet **homogeneous**.
- **Problem:** This is **not true** in reality. Moreover, not **control on cost** for user.
- **Current heuristics:**

$$EIpu(x) \triangleq \frac{EI(x)}{c(x)}$$

$$EI^{k}_{\text{cool}}(x) \triangleq \frac{EI(x)}{c(x)^{\tau_k}}$$

↪ Yet, **limited experimental performance** and **no theoretical justifications**.

A Pareto-Front intuition



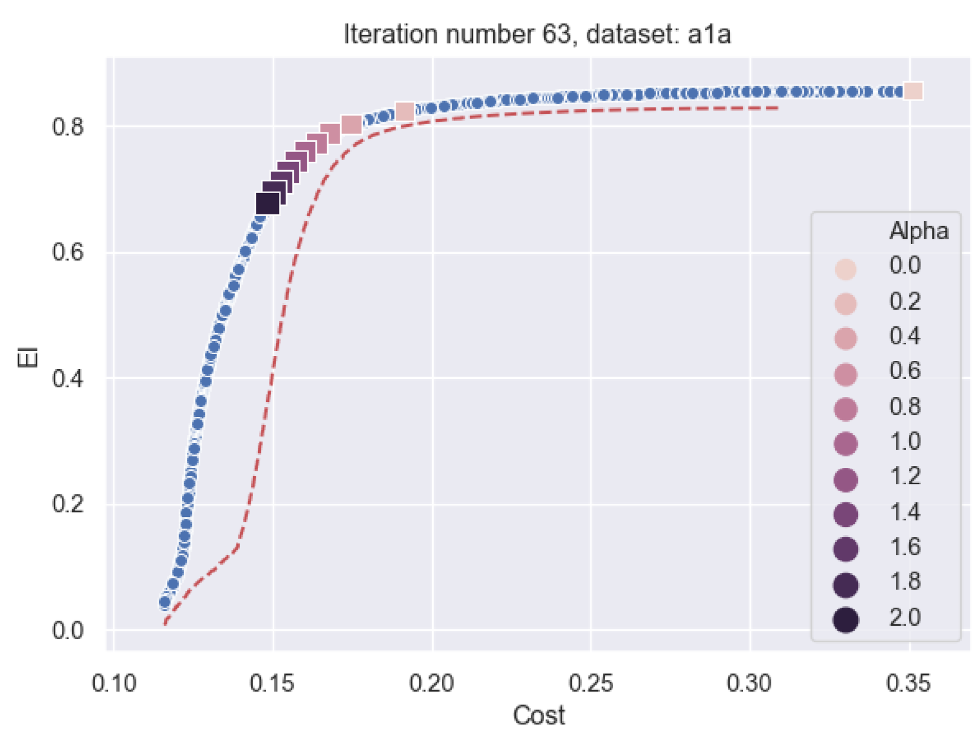


Figure: Representative examples of the EI-cost Pareto front at two different BO iterations. Yhe blue dots represent the Pareto front (EI,Cost) at the current BO iteration t, while there dashed curve refers to the Pareto front at iteration $t-1$.

- Towards a **better understanding of previous heuristics.**
- **Findings to leverage:** Consistent and general functional form (and unpredictable evolution, lack of optimality persistence)
- In theory and practice, two different settings:
  - **Bi-Optimization Problem**: Find an **optimal trade-off between the accuracy and cost** for given iteration budget.
  - **Optimal Time Allocation Problem**: Maximize accuracy under a cost budget constraint (no more restrictions on iterations number).

Pareto Efficient Expected Improvement

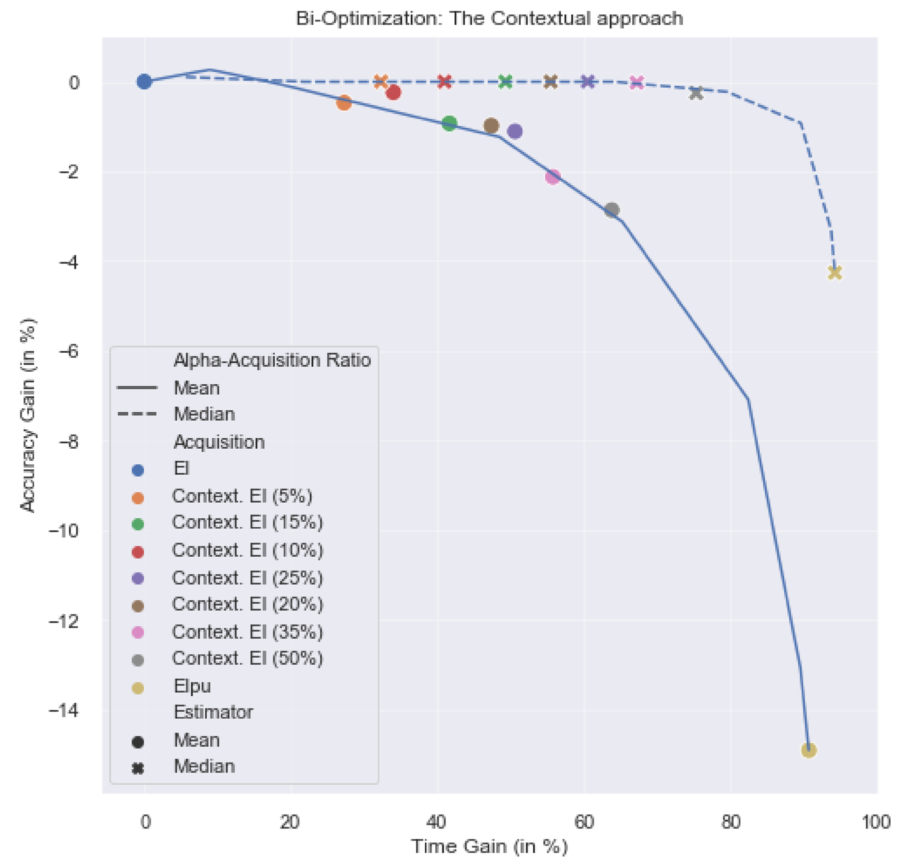


Figure: Bi-Optimization: (Blue curve) The accuracy-cost trade-off for EI_α at a set of α levels (i.e. you can trade-off x% accuracy for y% time gain)(Colored Points:) The accuracy-cost trade off for CEI_λ at a set of λ levels.

- Step 1: **Parametric generalization**

$$EI_\alpha(x) = \frac{EI(x)}{c(x)^\alpha},\ \alpha\in\mathbb{R}^+.$$

↪ We can **control and predict the cost-accuracy trade-off.**

- Step 2: **Dynamic allocation - Pareto Robustness**

$$\text{CEI}_\lambda(x) := \begin{cases} -c(x) & \text{if } EI(x) \geq (1-\lambda)\max_{z\in}(EI(z)), \\ -\infty & \text{otherwise.}\end{cases}$$

↪ Same **Performance** with more **Robustness**

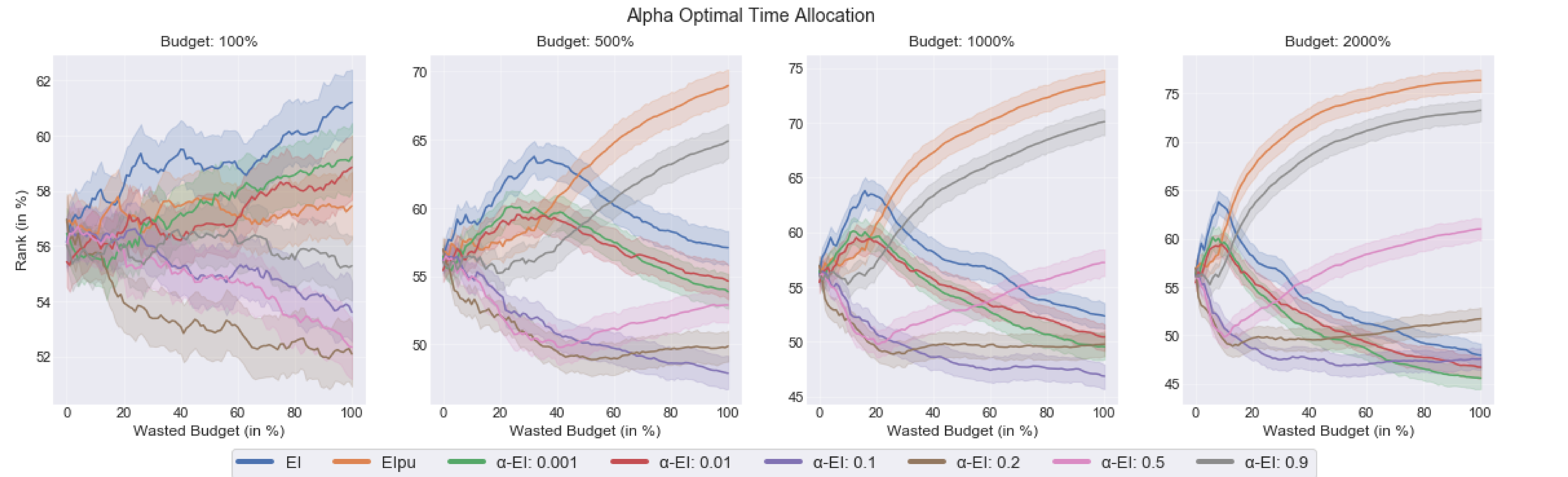


Figure: Comparison of EI_α with EI and EIpu in the optimal time allocation problem. Each plot corresponds to a different multiple of minimal budget (e.g. 2000% is 20 times this budget). Results are ranked at each iteration based on the minimum found up to that point by each method, a lower rank corresponding to a better minimization performance.

Cost Modeling: Online and Offline approaches

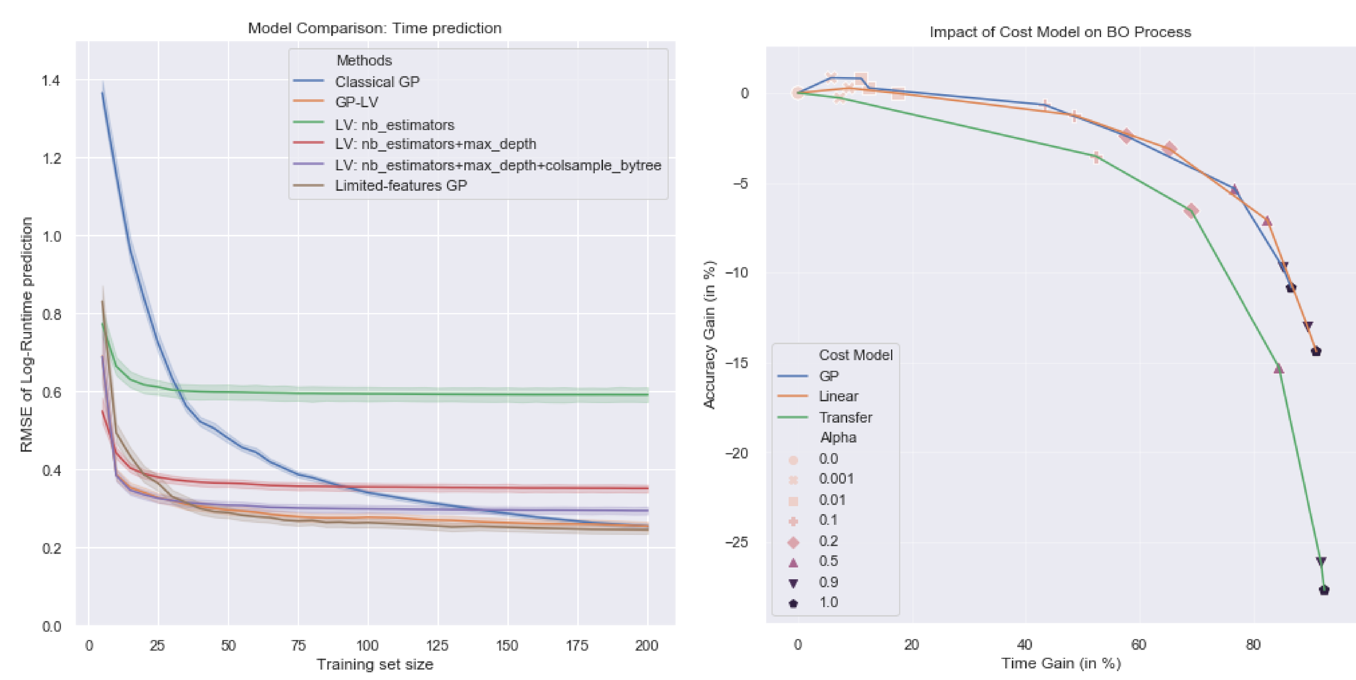


Figure: (Left): Performance of different cost models in grey-box setting. (Right): Impact of cost model on BO performance.

- **Online cost modelling**: Can we have better online cost modeling in grey box setting?
  ↪ Yes, with **low-variance models**, high importance of low-data regime.
- **Offline cost modelling**: Can we transfer offline cost models in grey box setting?
  ↪ Not Easy! Poor performance compared to simple online models.