UAI 2022

Combinatorial Bayesian Optimization with Random Mapping Functions to Convex Polytopes

Jungtaek Kim (POSTECH)   Seungjin Choi+ (Intellicode)   Minsu Cho+ (POSTECH)

+Co-Corresponding Authors

POSTECH POHANG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Introduction

- Bayesian optimization is a popular method for solving global optimization of a black-box function.
- When input variables are categorical or discrete, an extra care is needed.
- A common approach is to use one-hot encoded or Boolean representation for categorical variables which might yield a combinatorial explosion problem.
- We present a method for Bayesian optimization in a combinatorial space.

Contributions

- The main idea is to use a random mapping which embeds the combinatorial space into a convex polytope in a continuous space.
- All essential process is performed to determine a solution to the black-box optimization in the combinatorial space.
- The continuous space is defined as a convex polytope, a generalization of 0-1 polytope.
- Based on this perspective, we propose a combinatorial Bayesian optimization strategy with a random mapping function to a convex polytope and their lookup table, referred to as CBO-Lookup.
- Finally, we guarantee that our method has a sublinear cumulative regret bound, and demonstrate that our method shows satisfactory performance compared to other methods in various experiments.

Combinatorial Bayesian Optimization

- We consider the problem of minimizing a black-box function $f(\mathbf{c}) : \mathcal{C} \to \mathbb{R}$, where the input $\mathbf{c} = [c_1, \ldots, c_k]^\top$ is a collection of k categorical (or discrete) variables with each c_i taking one of N_i distinct values:

$$\mathbf{c}^\star = \underset{\mathbf{c} \in \mathcal{C}}{\arg\min}\, f(\mathbf{c}), \tag{1}$$

where the cardinality of the combinatorial space $\mathcal{C}$ is $|\mathcal{C}| = N_1 \times \cdots \times N_k = N$.
- Using Bayesian optimization, a global solution $\mathbf{c}^\star$ is found by repeating the construction of a surrogate and the optimization of an acquisition function.
- It estimates a surrogate $\widehat{f}$ to determine a candidate of global solution $\mathbf{c}^\dagger$,

$$\mathbf{c}^\dagger = \underset{\mathbf{c} \in \mathcal{C}}{\arg\max}\, a(\mathbf{c}; \widehat{f}(\mathbf{c}; \mathbf{C}, \mathbf{y})). \tag{2}$$

- This formulation has two technical challenges in (i) modeling a surrogate function over $\mathbf{c}$ and (ii) optimizing an acquisition function over $\mathbf{c}$.

Regrets

- Instantaneous regret r_t is a discrepancy between function values over $\mathbf{c}_t^\dagger$ and $\mathbf{c}^\star$:

$$r_t = f(\mathbf{c}_t^\dagger) - f(\mathbf{c}^\star). \tag{3}$$

- Using (3), we define a cumulative regret:

$$R_T = \sum_{t=1}^{T} r_t. \tag{4}$$

- Convergence is validated to prove the cumulative regret is sublinear, $\lim_{T\to\infty} R_T/T = 0$.

Algorithm 1 Combinatorial Bayesian Optimization with Random Mapping Functions

Initialize $\mathbf{C}_1 \subset \mathcal{C}$.
Update $\mathbf{y}_1$ by observing $\mathbf{C}_1$.
for $t = 1, \ldots T$ **do**
  Estimate a surrogate function $\widehat{f}(\mathbf{x}; \phi, \mathbf{C}_t, \mathbf{y}_t)$.
  Acquire a query point:

$$\mathbf{x}_t^\dagger = \underset{\mathbf{x} \in \tilde{\mathcal{X}} \subseteq \mathcal{X}}{\arg\max}\, \tilde{a}(\mathbf{x}; \widehat{f}(\mathbf{x}; \phi, \mathbf{C}_t, \mathbf{y}_t)). \tag{5}$$

  Recover $\mathbf{x}_t^\dagger$ to $\mathbf{c}_t^\dagger$.
  Observe $\mathbf{c}_t^\dagger$.
  Update $\mathbf{C}_t$ to $\mathbf{C}_{t+1}$ and $\mathbf{y}_t$ to $\mathbf{y}_{t+1}$.
end for
Find the best point and its function value:

$$(\mathbf{c}_{\text{best}}, y_{\text{best}}) = \underset{(\mathbf{c}, y) \in (\mathbf{C}_T, \mathbf{y}_T)}{\arg\min}\, y. \tag{6}$$

return $\mathbf{c}_{\text{best}}$

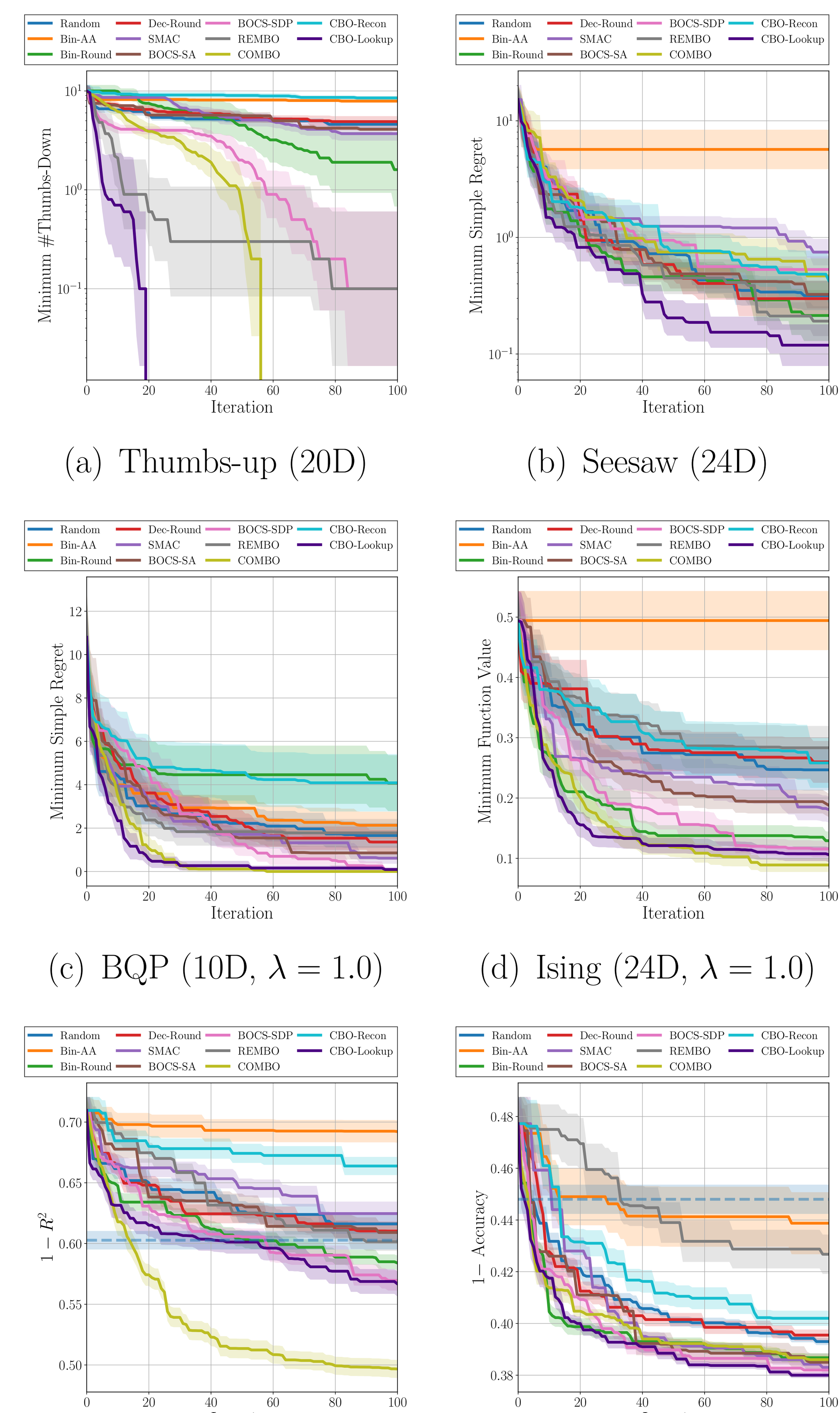
Lookup Tables

- In particular, we design a mapping function with a lookup table $\mathbf{L}$, constructed by a uniformly random matrix $\mathbf{R}$.
- Each row of the table is a key and value pair, where key and value indicate a combination and its embedding vector.
- After determining a query point, we find the closest point in $\mathbf{L}$ and then recover to one of possible combinations on $\mathcal{C}$.

Theorem 1. *Let $\delta \in (0,1)$, $\varepsilon \in (0,1)$, and $\beta_t \in \mathcal{O}(\sqrt{\log(\delta^{-1}Nt)})$. Suppose that a function f is on a RKHS $\mathcal{H}_\mathcal{C}$ and a kernel k is bounded, $k(\cdot,\cdot) \in [0, k_{max}]$. Then, the kernel-agnostic cumulative regret of combinatorial Bayesian optimization is upper-bounded with a probability at least $1-\delta$:*

$$R_T \in \mathcal{O}\left(\sqrt{(\sigma_n^2 + \varepsilon^2) NT \log(\delta^{-1}NT)}\right), \tag{7}$$

where σ_n^2 is the variance of observation noise and N is the cardinality of $\mathcal{C}$.



(a) Thumbs-up (20D)
(b) Seesaw (24D)
(c) BQP (10D, $\lambda = 1.0$)
(d) Ising (24D, $\lambda = 1.0$)
(e) Sparse regression
(f) Sparse classification

Figure 1: Experimental results on various circumstances.

Available on arXiv

arXiv

Contact Information

- Homepage: jungtaek.github.io
- GitHub: github.com/jungtaekkim