ClassBO: Bayesian Optimization for Heterogeneous Functions

Mohit Malu[1,3](✉), Giulia Pedrielli[2], Gautam Dasarathy[1], and Andreas Spanias[1,3]

[1] School of ECEE, Arizona State University, Tempe, AZ 85281, USA
{mmalu,gautamd,spanias}@asu.edu
[2] SCAI, Arizona State University, Tempe, AZ 85281, USA
giulia.pedrielli@asu.edu
[3] SenSIP Center, Tempe, USA

Abstract. Bayesian Optimization (BO) frameworks typically assume the function to be optimized is stationary (homogeneous) over the domain. However, in many real-world applications, we often deal with functions that present a rate of variation across the input space. In this paper, we optimize functions where a finite set of homogeneous functions defined over partitions of the input space can represent the heterogeneity. The disconnected partitions that can be characterized by the same function are said to be in the same class, and evaluating the function at input returns the minimum distance to a boundary of the contiguous class (partition). The ClassGP modeling framework, previously developed to model for such heterogenous functions along with a novel ClassUCB acquisition function and partition sampling strategy, is used to introduce a novel tree-based optimization framework dubbed as ClassBO (Class Bayesian Optimization). We demonstrate the superior performance of ClassBO against other methods via empirical evaluations.

Keywords: Bayesian Optimization · Gaussian process · Black-box Optimization · Heterogeneous function · Non-stationary function

1 Introduction

Bayesian optimization (BO) has emerged as a powerful sequential optimization approach for non-convex expensive to evaluate black-box functions [1,3]. The standard BO typically assumes the function to be optimized is stationary (homogeneous) over the domain, which allows using a single covariance kernel function with constant hyperparameters over the entire domain to model the function accurately. However, in many emerging applications such as machine learning, neural networks, and cyber-physical systems, the function to be optimized is heterogeneous, and the standard BO framework cannot accurately model these functions, leading to inadequate optimization performance. Often, heterogeneous functions can be characterized by locally stationary and globally non-stationary

P. Festa et al. (Eds.): LION 2024, LNCS 14990, pp. 249–253, 2025.
https://doi.org/10.1007/978-3-031-75623-8_19

functions, calling for more sophisticated optimization techniques that utilize the underlying structure to model and optimize these functions.

Many approaches have been proposed to extend BO for heterogeneous function optimization. Bayesian treed Gaussian Process (GP) in [5] and TuRBO in [2] use a collection of locally stationary GP's to model and optimize heterogeneous functions. Various methods that use non-stationary kernels have been proposed [4,8]. Methods in [7,9] warp the input space to a new space where the function is stationary and the standard BO framework can be applied. Compared to these methods, our approach utilizes the underlying structure, which allows the sharing of information across non-contiguous partitions if they share the same function, i.e., they are in the same class [6]. Specifically, we adopt the modeling architecture presented in [6] to develop a novel tree-based optimization algorithm.

Our contributions include: (i) A novel ClassBO framework to optimize heterogeneous functions; (ii) A set of novel ClassBO acquisition functions; (iii) Empirical analysis of ClassBO with different acquisition functions and compare it against other optimization techniques.

2 Notations and Problem Setup

We want to find the optimum of a heterogeneous function $f : \mathcal{X} \subseteq \mathbb{R}^d \rightarrow \mathbb{R}$ where the heterogeneity can be represented by a finite set of homogeneous functions g_j's defined over the axis-aligned partitions of the input space. Further, multiple partitions associated with the same function belong to the same class. For many engineering systems, the structure of the non-stationarity is known or can be evaluated. Hence, the observation model is such that evaluating the function at any point $\mathbf{x}$ reveals function evaluation (y), the class label (z), and the minimum distance to a boundary of partition (w). The optimization problem is formally given as follows:

$$\arg\max_{\mathbf{x}\in\mathcal{X}} f(\mathbf{x}) = \arg\max_{\mathbf{x}\in\mathcal{X}} \left(\sum_{j=1}^{p} \mathbb{1}\{\mathbf{x} \in \mathcal{X}_j\} g_j(\mathbf{x}) \right) \tag{1}$$

3 The ClassBO Algorithm

The ClassBO algorithm has three key components - (i) Bayesian statistical model: the statistical modeling framework, ClassGP, introduced in [6] is used to model heterogeneous functions by formulating a distribution over the space of objectives and computing the posterior conditioned on the observed samples; (ii) Acquisition functions: A set of novel ClassBO acquisition functions described in Section 3 are used to navigate the input space efficiently; (iii) Partition sampling strategy: A novel sampling strategy to learn the partitions of the input space accurately.

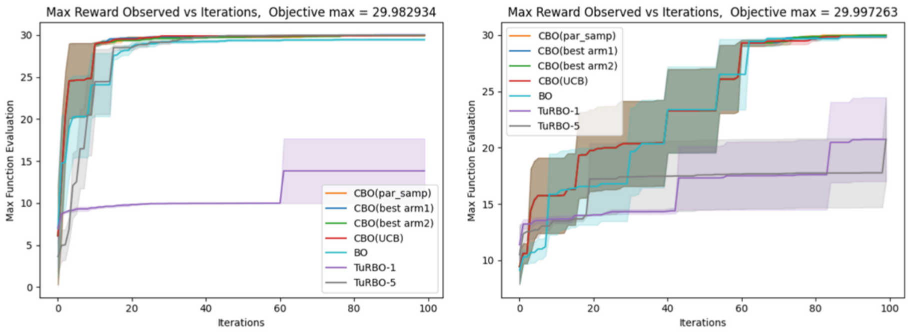
ClassUCB Acquisition Functions: BO algorithms use the acquisition function to decide where to sample in the next iteration, as the acquisition function can quantify the potential of finding an objective maximum at any given point in the input space. In this work, we formulate a set of novel acquisition functions for the ClassBO algorithm that uses mean and uncertainty estimates of the posterior distribution of the functions in each partition. One of the ClassUCB acquisition functions is given as follows:

$$(\mathrm{CBO-UCB}) \quad \mathbf{x}_t = \underset{\mathbf{x}\in\mathcal{X}_j, j\in[p]}{\arg\max} \left(\mu_{j,t_j}(\mathbf{x}) + \beta_{t_j}^{1/2}\sigma_{j,t_j}(\mathbf{x})\right)$$

Here, for the j^{th} partition learned using tailored CART algorithm of ClassGP, μ_{j,t_j}, σ_{j,t_j} represent the posterior mean and variance respectively of the function being modeled in the given partition after t_j iterations, β_{t_j} is a parameter that controls the trade-off between exploration and exploitation, and t is the current iteration of the algorithm. The ClassUCB acquisition function forms a set of points that maximize the UCB in each partition $j \in [p]$ and selects the point with maximum UCB from the set as the next sampling point.

Partition Sampling Strategy: ClassBO performs poorly when the partitions of the input space learned by the tailored CART algorithm are inaccurate or not complete. To resolve this, we run a proposed novel partition sampling strategy before using ClassUCB acquisition functions for sampling. The partition sampling strategy is a bottom-up approach applied to each partition learned by the tailored CART algorithm applied to initial samples. This approach constructs an axis-aligned hyper-rectangle for every sampled point of the length twice the minimum distance from the boundary (w) with the sampled point at the center and uniformly samples from the region not covered by the hyper-rectangles. This approach guarantees all the partitions of input space are learned accurately. The pseudo-code for the algorithm is given as follows:

Step 1: For each sampled point *construct* an axis-aligned hyper-rectangle of length twice the minimum distance from the boundary with the sampled point at the center.
Step 2: *Merge* hyper-rectangles that overlap or are sufficiently close within a given partition to form a larger hyper-rectangle.
Step 3: Uniformly *sample* from the regions that is not covered by the hyper-rectangles.
Step 4: *Stop* - If the final merged hyper-rectangle covers the entire partition else *repeat*.



(a) $p = 4$ (b) $p = 16$.

Fig. 1. Maximum observed reward vs iterations to compare of performance of ClassBO with ClassUCB acquisition functions and other baselines for fixed parameters. {number of classes ($k = 2$), initial samples ($n = 5$), iterations ($T = 100$)}

4 Simulation Results

Simulations were performed over the function constructed using standard optimization test functions. We plot the maximum observed reward (averaged over multiple runs) vs iterations to compare the performance. Following parameters are initialized for each simulation: dimension ($d = 2$), initial samples ($n = 5$), number of partitions ($p = 4, 16$), number of classes ($k = 2$), number of iterations ($T = 100$), and for a fixed set of initialized parameters 5 independent simulation runs for each framework are performed for comparison. The results in Fig. 1(a) shows that ClassBO outperforms BO and TuRBO algorithms. However, as the number of partitions increase, the partitions sampling strategy ends up sampling regions of lower interest to learn the partitions accurately, in turn leading to slower convergence to the optima as observed in Fig. 1(b).

5 Conclusions and Future Work

In this paper, we propose a new tree-based ClassBO framework that uses ClassGP modeling technique for heterogeneous functions with access to class information. We also introduce a set of novel ClassUCB acquisition functions and compare the performance of ClassBO against other baselines. Additionally, we establish that the performance of ClassBO is heavily dependent on the accuracy of learning the partition that contains the maximum of the objective function. For future work, improving the sampling strategy by incorporating uncertainties pertaining to learned partitions instead of deterministic partition sampling strategy, scaling to higher dimensions, and theoretical analysis of the algorithm are promising avenues to explore.

References

1. Archetti, F., Candelieri, A.: Bayesian Optimization and Data Science, vol. 849. Springer, Cham (2019)
2. Eriksson, D., Pearce, M., Gardner, J., Turner, R.D., Poloczek, M.: Scalable global optimization via local Bayesian optimization. In: Advances in Neural Information Processing Systems, vol. 32 (2019)
3. Frazier, P.I.: A tutorial on Bayesian optimization. arXiv preprint arXiv:1807.02811 (2018)
4. Gibbs, M.N.: Bayesian Gaussian processes for regression and classification. Ph.D. thesis. Citeseer (1998)
5. Gramacy, R.B., Lee, H.K.H.: Bayesian treed gaussian process models with an application to computer modeling. J. Am. Stat. Assoc. **103**(483), 1119–1130 (2008)
6. Malu, M., Pedrielli, G., Dasarathy, G., Spanias, A.: Class GP: Gaussian process modeling for heterogeneous functions. In: International Conference on Learning and Intelligent Optimization, pp. 408–423. Springer (2023)
7. Marmin, S., Ginsbourger, D., Baccou, J., Liandrat, J.: Warped gaussian processes and derivative-based sequential designs for functions with heterogeneous variations. SIAM/ASA J. Uncertainty Quantification **6**(3), 991–1018 (2018)
8. Paciorek, C.J., Schervish, M.J.: Spatial modelling using a new class of nonstationary covariance functions. Environmetrics: Official J. Int. Environmetrics Soc. **17**(5), 483–506 (2006)
9. Schmidt, A.M., O'Hagan, A.: Bayesian inference for non-stationary spatial covariance structure via spatial deformations. J. R. Stat. Soc.: Series B (Stat. Methodol.) **65**(3), 743–758 (2003)