Proceedings of the 2025 Winter Simulation Conference
E. Azar, A. Djanatliev, A. Harper, C. Kogler, V. Ramamohan, A. Anagnostou, and S. J. E. Taylor, eds.

PRIOR-DATA FITTED NETWORKS FOR MIXED-VARIABLE BAYESIAN OPTIMIZATION

Timothy Shinners[1]

[1]Dept. of Computer Science, Universität der Bundeswehr München, Munich, GERMANY

ABSTRACT

Bayesian optimization is an algorithm used for black-box optimization problems, such as parameter tuning for simulations. Prior-data fitted networks (PFNs) are transformers that are trained to behave similarly to Gaussian processes (GPs). They have been shown to perform well as surrogate functions in Bayesian optimization methods, offering performance capabilities similar to GPs with reduced computational expense. PFNs have not yet been applied to settings with a mixed-variable input space that involves both numerical and categorical variables. In this work, we train three new PFNs using existing mixed-variable GPs. We integrate them into mixed-variable Bayesian optimization (MVBO) methods and conduct experiments with six different black-box functions to assess their behavior. Our PFNs yield comparable quality of performance to that of their GP-based counterparts in MVBO settings, while operating at drastically reduced computational expense.

1 INTRODUCTION

Simulation-based optimization can be posed as a black-box optimization problem, in which a candidate solution is selected, the simulation is run, and a result is observed. Based on the set of previous observations, a new candidate is selected, and the process is repeated. Bayesian optimization is an algorithm that aims to increase the efficiency of solving black-box problems by optimizing the candidate suggestions. First, a surrogate model, usually a Gaussian process, is fit to the set of previous observations. Using the surrogate model's predictions and uncertainty, it is possible to compute an acquisition function, such as the "expected improvement" (Mockus et al. 1978). The acquisition function estimates the utility of evaluating a given candidate, providing a balance between exploration and exploitation. The acquisition function is cheap to compute, so it is optimized via gradient descent to produce a suggestion for the next iteration.

Often, real-world black-box optimization problems involve categorical inputs. These problems require mixed-variable Bayesian optimization methods that are designed to handle both categorical and numerical variables. CoCaBO (Ru et al. 2019), Casmopolitan (Wan et al. 2021), and BODi (Deshwal et al. 2023) are all proposed methods for mixed-variable Bayesian optimization. They employ Gaussian processes that mix numerical and categorical kernels.

Prior-Data Fitted Networks (PFNs) are transformers that have been trained to perform Bayesian inference (Müller et al. 2022). As input, they take an entire training data set, and the inputs from a test set. For each observation in the test set, the PFN outputs a predictive posterior distribution over the target. Calculating posterior distributions for a Gaussian process requires $\mathscr{O}(n^3)$ time for an amount of data n, so they do not scale well with large amounts of data. PFNs perform the same task with magnitudes greater efficiency.

PFNs are trained using synthetic data generated from a prior, usually a Gaussian process. Entire data sets are sampled from this prior, which are split into training and test sets. As input, the PFN takes the training set and the test inputs, and outputs a Riemann distribution over the test targets, which are discretized approximations of continuous probability distributions.

Müller et al. (2023) showed that PFNs could be used as a surrogate model in Bayesian optimization. They performed competitively against the GPs they were trained on, at a fraction of the computational expense. However, PFNs were not applied to settings with mixed-variable input spaces.

2 CONTRIBUTIONS

We trained three new PFNs, focusing on mixed-variable settings. The Gaussian processes from CoCaBO (Ru et al. 2019), Casmopolitan (Wan et al. 2021), and BODi (Deshwal et al. 2023) were used as priors.

3 EXPERIMENTS AND RESULTS

To test the performance of different BO methods, five synthetic functions and one real-world task were selected for experimentation. Each optimization trial was initialized with 10 randomly drawn observations, and 30 trials were completed for each task.

As seen in Figure 1, the new PFNs are very competitive with their GP counterparts in mixed-variable settings. For long optimization runs with more than 500 iterations, PFNs perform with drastically reduced computational expense. Results vary depending on the given task, but overall, PFNs offer an appealing alternative to GPs as a surrogate function in mixed-variable Bayesian optimization methods.

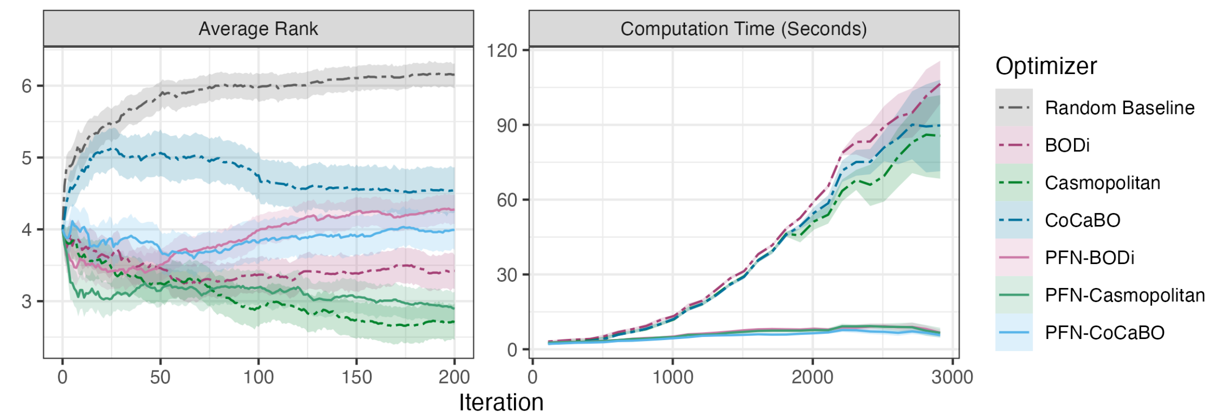

Figure 1: The left plot shows each optimizer's rank, at each BO iteration, averaged across all trials. Ranks were determined by the best value obtained by each optimizer up to that iteration. The right plot shows the time taken per iteration on extended optimization runs. Shaded regions represent 95% confidence intervals.

REFERENCES

Deshwal, A., S. Ament, M. Balandat, E. Bakshy, J. R. Doppa, and D. Eriksson. 2023. "Bayesian Optimization over High-Dimensional Combinatorial Spaces via Dictionary-based Embeddings". *CoRR* abs/2303.01774 https://doi.org/10.48550/ARXIV.2303.01774.

Mockus, J., V. Tiesis, and A. Zilinskas. 1978. "The Application of Bayesian Methods for Seeking the Extremum". *Towards Global Optimization* 2(117-129):2.

Müller, S., N. Hollmann, S. P. Arango, J. Grabocka, and F. Hutter. 2022. "Transformers Can Do Bayesian Inference". In *International Conference on Learning Representations*.

Müller, S., M. Feurer, N. Hollmann, and F. Hutter. 2023. "PFNs4BO: In-Context Learning for Bayesian Optimization".

Ru, B., A. S. Alvi, V. Nguyen, M. A. Osborne, and S. J. Roberts. 2019. "Bayesian Optimisation over Multiple Continuous and Categorical Inputs".

Wan, X., V. Nguyen, H. Ha, B. Ru, C. Lu, and M. A. Osborne. 2021. "Think Global and Act Local: Bayesian Optimisation over High-Dimensional Categorical and Mixed Search Spaces".