Evaluated System for a Bayesian Optimization Service

Ian Dewancker1, Michael McCourt1, Scott Clark1, Patrick Hayes1, Alexandra Johnson1, and George Ke2

1SigOpt, 2University of Waterloo

Objectives

An evaluation framework for rigorously testing and quickly understanding the impact of changes to the SigOpt optimization service.

- Perform end-to-end testing of service
- Compare between algorithm versions
- Compare against external baselines

Introduction

SigOpt offers an optimization service to help customers tune complex systems, simulations and models. Our optimization engine applies several concepts from Bayesian optimization [1] and machine learning to optimize customers metrics as quickly as possible.

In particular, we consider problems where the maximum is sought for an expensive function $f: \mathcal{X} \rightarrow \mathbb{R}$.

$$x_{\text{opt}} = \arg \max_{x \in \mathcal{X}} f(x)$$

SigOpt’s core optimization engine is a closed-source fork of the open-source MOE project [2]. The SigOpt service supports a succinct set of HTTP API endpoints for optimizing objective functions.

Metrics

The performance metrics we consider for comparisons on a given objective function are the best value seen by the end of the optimization (Best Found), and the area under the best seen curve (AUC). The AUC metric can help to better differentiate performance, as shown in Figure 1. Optimization algorithms are run at least 20 times on each function. The distributions of the performance metrics are compared using the non-parametric Mann-Whitney U test, used in previous studies of Bayesian optimization methods [3] [4].

Benchmark Suite

The tests for our evaluation system consist of closed-form optimization functions [5] which are extensions of an earlier set proposed for black-box optimization evaluation. The benchmark suite is open source.

Infrastructure

Lightweight function evaluation processes are run concurrently on a master machine with many cores. External baseline methods including TPE, SMAC, Spearmint, particle swarm optimization and random search are run locally on master.

For inspecting comparative performance on individual functions, we plot the best value of the objective metric seen after each function evaluation.

Visualization Tools

To better summarize the relative performance between two methods $(A, B)$ using metric $M$, we create histograms of the $p$-values returned by the Mann-Whitney U tests. Test functions are first partitioned by the relative expected value of the metric $M$, then histograms are generated for the two sets.

$$\text{wins}(A > B)_M = \{ \text{func} \mid E[M_A] > E[M_B] \}$$

$$\text{wins}(B > A)_M = \{ \text{func} \mid E[M_B] > E[M_A] \}$$

Conclusions

Our evaluation system has been become a valuable analysis tool when considering algorithm or system changes to the SigOpt optimization service. Data driven performance analysis is an effective way to enable faster iteration and evaluation of a wide spectrum of ideas. The system continues to guide improvements to the core SigOpt service by providing empirical comparisons between internal changes, alternative optimization methods, as well helping to expose errors and bugs.

References


Contact Information

- Web: http://www.sigopt.com
- Email: contact@sigopt.com