Proceedings of the 2017 Winter Simulation Conference
W. K. V. Chan, A. D'Ambrogio, G. Zacharewicz, N. Mustafee, G. Wainer, and E. Page, eds.

ACQUISITION FUNCTIONS FOR SIMULTANEOUS BAYESIAN OPTIMISATION OF MULTIPLE PROBLEMS

Michael Pearce

Centre for Complexity Science
Zeeman Building, Warwick University
Coventry, UK, CV4 7AL

ABSTRACT

Using Gaussian Processes as statistical predictors of expensive objective functions for optimization has gathered much attention over the last two decades. Many acquisition functions that guide the search for data collection have been developed for various optimization cases, including multi-fidelity and multiple objectives. We look at the case where there are multiple tasks to be solved simultaneously. One example may be the algorithm selection problem where each use case of an algorithm is a unique optimization problem, and a user aims to find the optimal setting for each use case based on optimal settings of similar cases. Input uncertainty can be seen as a related problem where one must use the same settings for all use cases. We propose a variety of acquisition functions for Bayesian optimization in this general class of optimization problems.

1 INTRODUCTION

Bayesian Optimization is the field of using cheap statistical surrogate models to aid the optimization of expensive to evaluate functions. Gaussian Processes are the most commonly used surrogate model. Due to their Bayesian nature, a user can aid the optimization procedure with abstract prior knowledge, and due to their mathematical simplicity, for a given input the posterior distribution over possible function values has a closed form, under certain assumptions. The posterior distribution can be used to select promising new points of the expensive function to evaluate, and the expected value of a new evaluation is given by the so-called acquisition function. There exist many acquisition functions. The Efficient Global Optimization (EGO) algorithm uses a Gaussian Process with a simple measure of the statistically expected increase, if any, that a new real function value will have over best function value found so far. Knowledge Gradient (KG) collects function values that maximize the expected increase in the *predicted* peak of the model, not just the best of the evaluated points like EGO. Using the expected reduction in entropy over the distribution of the peak has also been widely used in the machine learning community.

In recent years, there has been interest in multi-task optimization where many optimization tasks exist and various objectives can be constructed. Unlike mutli-objective optimization where each new function evaluation yields multiple outputs, for multi-task optimization, a user, or an acquisition function, must select both a task and a tool, and then a function evaluation yields a single scalar output. Three large areas of research may be considered as multi-task optimization. Firstly the aim of the user may be to find the best tool for each task, for example, finding the best algorithm for each use case, finding the best treatment for each patient, finding the best action for every state in a control problem. Secondly, a user may aim to find the single best tool averaged over all tasks, which is relevant in the case of input uncertainty. Thirdly, a user may aim to find the best input on a large slow to evaluate task by using information from other quicker cheaper tasks, i.e., multi-fidelity optimization. Much work has been done on multi-fidelity optimization therefore we focus on the first two cases, simultaneous optimization of multiple tasks and optimization under task uncertainty. We have adapted and extended the Knowledge Gradient algorithm to both of these cases and the popular EGO algorithm to the task uncertainty case. Where alternative algorithms exist we have compared and outperformed such methods by exploiting similarity across tasks.

1 METHODS

The fundamental design of the EGO algorithm is that at the end of sampling the algorithm will return the function input with the best associated evaluation from the dataset to the user. Therefore the current best point within the dataset is used as a point of comparison and the expectation of the possible increase of a new point is used as the acquisition function.

Adapting this to the case of task uncertainty we simply need to answer the following questions, how do we determine the "best" input over tasks to return to the user, and how do we measure improvement in this input caused by a single new evaluation from an input on a single task. We simply answer these questions by replacing the value of the largest data point in the single task case with the average predicted value of a Monte-Carlo integral over the uncertain tasks. The influence of a new data point on one (tool,task) combination affects the Gaussian Process also for other tasks. By measuring this influence using Bayesian adaptive updating formula, one can evaluate how much the Monte-Carlo integral of a new input will improve over the current best Monte-Carlo integral caused by only one evaluation of a tool on one task. This modification process can also be easily applied to the Knowledge Gradient algorithm, each input has an associated Monte-Carlo integral, and the expected increase in the largest of the integrals can be measured by the single-task Knowledge Gradient acquisition function.

For the case of simultaneous multi-task optimization, we assume that a user aims to maximise a weighted sum over all the task optima, with user-defined weights. Executing one tool on one task will yield an evaluation that updates the model and affects the optima of all other tasks. The Knowledge Gradient algorithm is used to measure the expected increase in each optimum, and by taking the weighted sum we derive the Regional Expected Value of Improvement (REVI) acquisition function. By sampling according to the REVI algorithm we obtain much faster convergence to the optima of all tasks compared to any other known method from the literature.

REFERENCES

Scott, W., P. Frazier, and W. Powell. 2011. "The Correlated Knowledge Gradient for Simulation Optimization of Continuous Parameters using Gaussian Process Regression". *SIAM Journal on Optimization* 21 (3): 996–1026.

Jones, D. R., M. Schonlau, and W. J. Welch. 1998. "Efficient Global Optimization of Expensive Black-Box Functions". *Journal of Global Optimization* 13 (4): 455–492.