Recent Advances in Bayesian Optimization

Dr Vu Nguyen
vu@robots.ox.ac.uk
University of Oxford, UK

Abstract

Bayesian optimization (BO) has emerged as an exciting sub-field of machine learning and artificial intelligence that is concerned with optimization using probabilistic methods. Systems implementing BO techniques have been successfully used to solve difficult problems in a diverse set of applications, including automatic tuning of machine learning algorithms, experimental designs, and many other systems. Several recent advances in the methodologies and theory underlying BO have extended the framework to new applications and provided greater insights into the behavior of these algorithms. Bayesian optimization is now increasingly being used in industrial settings, providing new and interesting challenges that require new algorithms and theoretical insights. Therefore, I think having a tutorial on Bayesian optimization for ACML audience is timely, useful, and practical for both academia and industries to know the recent advances on Bayesian optimization in a systematic manner. The topics of this tutorial consists of two main parts. In the first part, I will go into detail the BO in the standard setting. In the second part, I will present the current advances in Bayesian optimization including (1) batch BO, (2) high dimensional BO and (3) mixed categorical-continuous BO. In the end of the talk, I also outline the possible future research directions in Bayesian optimization.

Tutorial Website: https://ntienvu.github.io/BayesianOptimizationTutorial_ACML2020.html

1 Goals and Objectives

Specific goals and objectives The goal of this tutorial is offering ACML audience a timely, useful, and practical introduction of Bayesian optimization in a systematic manner. In particular, my aim is to introduce the techniques, applications and future research directions of BO. In addition, I provide a broad summary of recent advances in batch BO, in unknown search space and in high dimension settings.

Why is the topic important/interesting to ACML audience? There is proliferation of machine learning and data mining algorithms which hyper-parameters tuning is needed. Grid search and random search are two popular approaches for finding the best hyper-parameters. Can we do better than grid search and random search? This tutorial provides insides on how BO can be used to tune the model parameters in the fewest iteration with theoretical guarantee. Bayesian optimization will lift up the performances of all machine learning and data mining algorithms which are sensitive to the choice of hyper-parameters. *As a result, all ACML audience who are dealing with hyper-parameter tuning for their algorithms will be greatly benefited.*

What is the expected background of the audience? I do not require the audiences to have strong background knowledge on Bayesian modeling. However, we expect the audience already understand some basic concepts and terminologies on artificial intelligence, data mining, and machine learning.

Description of the previous versions of tutorial

- A shorter version of this tutorial has been delivered at University of Twente (The Netherland), University of Liverpool (UK), University of Glasgow (UK) in June 2020, Amazon Research in July 2020.
- The GP part and the demo for BO have been taught in class lectures by Dr Vu Nguyen at University of Oxford in Oct 2019.
- Different parts of the tutorial are presented by the speaker and his colleagues at ICDM 2016, NIPSW 2016, IJCAI 2017, ICML 2017, ACML 2017, ICDM 2017, ICDM 2018, NeurIPS 2018, ICDM 2019, ICML 2020 respectively.

2 Outline including a short summary of every section

2.1 Tutorial Outline and Motivation to Bayesian Optimization [5 min]

I will start with the problem of machine learning hyper-parameter tuning and experimental design which can be seen as the black-box function. Then, I will give a brief introduction to optimize these black-box functions using Bayesian optimization.

- Machine Learning Hyper-parameter Tuning and Experimental Design as Black-box Function
- Bayesian Optimization for Optimizing a Black-box Function

2.2 Part I. Bayesian Optimization [55 min]

Bayesian optimization is a sequential model-based approach to solving global optimization problem of black-box functions. By black-box function, we assume that the function f has no simple closed form, but can be evaluated at any arbitrary query point x in the domain. In particular, the BO framework has two key ingredients. The first ingredient is a probabilistic surrogate model, which consists of a prior distribution that captures our beliefs about the behavior of the unknown objective function and an observation model that describes the data generation mechanism. The second ingredient is a loss function that describes how optimal a sequence of queries are; in practice, these loss functions often take the form of regret, either simple or cumulative. Ideally, the expected loss is then minimized to select an optimal sequence of queries. After observing the output of each query of the objective, the prior is updated to produce a more informative posterior distribution over the space of objective functions.

- Bayesian Optimization [10 mins]
- Gaussian Processes and acquisition function [5 mins]
- Illustration of Bayesian Optimization [5 mins]
- Convergence Analysis in BO [5 mins]
- Applications of Bayesian Optimization [10 mins]
- Question and Answer [10 mins]

2.3 Part II. Recent Advances in Bayesian Optimization

2.3.1 Batch Bayesian Optimization [15 min]

Standard BO approaches allows the exploration of the parameter space to occur sequentially. Often, it is desirable to simultaneously propose batches of parameter values to explore. This is particularly the case when large parallel processing facilities are available. These could either be computational or physical facets of the process being optimized. Batch methods, however, require the modeling of the interaction between the different evaluations in the batch, which can be expensive in complex scenarios. In this section, I will summarize the recent batch BO models. I will provide the strengths and weaknesses of each approach.

- Introduction and Problem Statements [3 mins]
- Peak Suppression Approaches [3 mins]
- Budgeted Batch BO for Unknown Batch Size [3 mins]
- Thompson Sampling for Batch BO [3 mins]
- Asynchronous Batch BO [3 mins]

2.3.2 High Dimensional Bayesian Optimization [15 min]

Existing BO is limited to about 10 dimensions. Scaling BO methods to handle functions in high dimension presents two main challenges. Firstly, the number of observations required by the GP grows exponentially as input dimensions increase. This implies more experimental evaluations are required, often expensive and infeasible in real applications. Secondly, global optimization for high dimensional acquisition functions is intrinsically a hard problem and can be prohibitively expensive to be feasible. I will discuss recent advances in Bayesian optimization techniques for high dimensional settings.

- Introduction and Problem Statements [3 min]
- Existing Approaches in High Dimensional BO [3 min]
- High dim BO with Elastic GP [3 min]
- High dim BO using Dropout [3 min]
- High dim BO using Local Optimization. [3 mins]

2.3.3 Mixed Categorical-Continuous Bayesian Optimization [20 min]

Real-world optimization problems are typically of mixed-variable nature, involving both continuous and categorical input variables. For example, tuning the hyperparameters of a deep neural network involves both continuous variables, e.g., learning rate and momentum, and categorical ones, e.g., optimizer types, activation type. Having a mixture of categorical and continuous variables presents unique challenges. If some inputs are categorical variables, as opposed to continuous, then the common assumption that the BO acquisition function is differentiable and continuous over the input space, which allows the acquisition function to be efficiently optimized, is no longer valid. I will discuss recent advances in BO for mixed categorical-continuous settings.

- Introduction and Problem Statements
- Existing Approaches in High Dimensional BO [5 min]
- Multi-armed Bandits [5 mins]
- Categorical-specific Continuous Optimization. [5 mins]
- Categorical Continuous Optimization. [5 mins]

2.4 Future Research Directions and Q&A [15 min]

- Future Research Directions [5 min]
- Question and Answer [10 min]

3 Presenter Biography

Dr Vu Nguyen is currently a Senior Research Associate at a Machine Learning Research Group at University of Oxford. He is working with Professor Michael Osborne and Professor Andrew Briggs on a machine learning project for tuning quantum devices using Bayesian optimization and deep reinforcement learning. Previously he was working as a Research Scientist at a Credit AI in Melbourne and was a postdoctoral researcher at Deakin University where he obtained his PhD in 2015. He published regularly at top venues in machine learning. He was the recipient of ACML 2016 best paper award, IEEE ICDM 2017 best papers and one of the 200 young researchers world-wide for attending Heidelberg Laureate Forum 2015. He gains expertise on Bayesian Machine Learning and Bayesian Optimization with 20 papers published in premier venues, including ICML, NeurIPS, ICDM, IJCAI, AISTATS and ACML.