Flexible, Parallelized, Bayesian Optimization for Large Datasets

Theodore Long

November 31st 2022

FABOLAS – Bayesian Hyperparameter Tuning

- **Fa**st **B**ayesian **O**ptimization for **La**rge Data**S**ets
- How do we optimize hyperparameters to achieve optimal training accuracy?
- Training a model for each hyperparameter setting is **expensive**
- Solution: test performance on **subsets of full dataset**
- FABOLAS combines this idea with Bayesian Optimization

Bayesian Optimization with Small Subsets

- Classic Bayesian Optimization seeks to maximize **objective function** f

$$\operatorname*{argmax}_{x \in \mathbb{R}^n} f(x)$$

where $x \in \mathbb{R}^n$ is the hyperparameter setting (learning rate, batch size, etc.)

- FABOLAS adds an additional parameter

$$s = {|D_{small}|}/{|D|}$$

where s is the size of the random data subset D_{small} as a fraction of the full dataset D

Bayesian Optimization with Small Subsets

- Our new objective function now looks like

$$f(x, s)$$

$$x \in \mathbb{R}^n, s \in (0, 1]$$

- Our goal is to find

$$\underset{x \in \mathbb{R}^n}{\operatorname{argmax}} f(x, s = 1)$$

- FABOLAS can use small values of s to do ‘cheap’ experiments that are correlated with performance at $s = 1$

Limitations of FABOLAS

- Limited number of datasets/models tested in original paper
- Current Implementation is not parallelized
  - Ray Tune provides an easy interface for parallelization
- FABOLAS uses a GP for modeling objective function and cost function
  - This becomes expensive with many datapoints/dimensions
  - Can other models provide faster/more precise estimates?
- FABOLAS is not fully Bayesian
  - Only chooses from configuration points previously evaluated in training

Goals for this Project

- Test FABOLAS with a wider range of models and datasets using open-source BO benchmarks (e.g. HPOBench)
- Parallelize using Ray Tune
- Implement other model backends
- (Time permitting) Attempt more fully Bayesian approach

Questions?