Meta-Learning Acquisition Functions for Transfer Learning in Bayesian Optimization

Chen Zhao

ICLR 2020, https://arxiv.org/abs/1904.02642

Motivations

Bayesian optimization (BO)'s data efficiency originates from a probabilistic surrogate model which is used to generalize over information from individual data points. This model is typically given by a Gaussian process (GP), whose well-calibrated uncertainty prediction allows for an informed exploration-exploitation trade-off during optimization. The exact manner of performing this trade-off, however,is left to be encoded in an acquisition function (AF).

To achieve optimal data-efficiency on new instances of previously seen tasks, however, it is crucial to incorporate the information obtained from these tasks into the optimization. Therefore, transfer learning is an important and active field of research. Indeed, in many practical applications, optimizations are repeated numerous times in similar settings, underlining the need for specialized optimizers. Examples include hyperparameter optimization which is repeatedly done for the same machine learning model on varying datasets or the optimization of control parameters for a given system with varying physical configurations

Global Black-Box Optimization

Find a global optimium of some unknown objective function $f : \mathbb{R}^d \supset \mathcal{D} \to \mathbb{R}$:

$$\mathbf{x}^* \in \arg\max_{\mathbf{x}\in\mathcal{D}} f(\mathbf{x}) \tag{1}$$

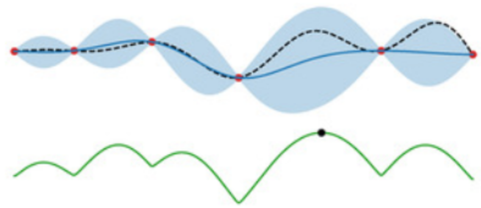
At step t, decide for $\mathbf{x}_t$ based exclusively on the optimization history $\mathcal{H} \equiv \{\mathbf{x}_i, \mathbf{y}_i\}_{i=1}^{t-1}$ with

$$\mathbf{y}_i \equiv f(\mathbf{x}_i) + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma_n^2) \tag{2}$$

Bayesian Optimization (BO)

- Data-efficient algorithm for global black-box optimization
- Probabilistic surrogate model (*e.g.* GP) to interpolate between data points

- Sampling strategy (acquisition function, AF) based on surrogate model
- Question: how to speed up BO with prior knowledge about f?



Transfer Learning for BO

- Increase data-efficiency by transferring knowledge across task instances $f_i \in \mathcal{F}$
- Function class $\mathcal{F}$ can be given, *e.g.* by
  - loss function of machine learning model evaluated on various datasets
  - varying physical configurations of laboratory experiment
- Source tasks can come from, *e.g.*
  - results from previous optimization runs
  - simulations

MetaBO

- Retain the proven structure of BO, keep the powerful GP surrogate model
- Transfer knowledge via "neural AFs": task-specific AFs represented as neural nets
- Train neural AFs using reinforcement learning, *i.e.*, no need for gradients of $f \in \mathcal{F}$