Statistica Sinica **36** (2026), 1-24
doi:https://doi.org/10.5705/ss.202023.0310

BAYESIAN OPTIMIZATION WITH PARETO-PRINCIPLED TRAINING FOR ECONOMICAL HYPERPARAMETER OPTIMIZATION

Yang Yang, Ke Deng* and Yu Zhu*

Nankai University, Tsinghua University and Purdue University

Abstract: The specification of hyperparameters plays a critical role in determining the practical performance of a machine learning method. Hyperparameter Optimization (HPO), i.e., the searching for optimal specification of hyperparameters, however, often faces critical computational challenges due to the vast searching space and the high computational cost on model training under a given hyperparameter specification. In this paper, we propose BOPT-HPO, a systematic approach for efficient HPO by leveraging Bayesian optimization with Pareto-principled training, based on the observation that the training procedure of a machine learning method under a given hyperparameter specification often follows the Pareto principle (the 80/20 rule) that about 80% of the total improvement in the objective function is achieved in 20% of the training time. By introducing two levels of training corresponding to the Pareto principle, i.e., the eighty-percent training (ET) and the complete training (CT), and establishing a joint surrogate model for CT runs and ET runs, BOPT-HPO reduces the computational cost of HPO significantly under the framework of Bayesian optimization with multi-fidelity measurements. A wide range of experimental studies confirm that the proposed approach achieves economical HPO for various machine learning models, including support vector machines, feed-forward neural networks, and convolutional neural networks.

Key words and phrases: Automated artificial intelligence, black-box function optimization, computer experiments, multi-fidelity modelling, truncated Gaussian process.

1. Introduction

Machine Learning (ML) models, such as Support Vector Machines (SVMs) and Deep Neural Networks (DNNs), have become a driving force for modern data-science technologies, leading to applications not only in artificial intelligence, such as computer vision (Krizhevsky, Sutskever and Hinton, 2012) and natural language processing (Goldberg, 2017), but also in intelligent manufacturing, such as material design (Batra, Song and Ramprasad, 2021) and 3D printing (Zhu et al., 2021).

In addition to the usual model parameters, a ML model typically involves hyperparameters that define the model's structure (e.g., the number of hidden

*Corresponding author. E-mail: (Deng) kdeng@tsinghua.edu.cn; (Zhu) yuzhu@purdue.edu