Proceedings of the 2021 Winter Simulation Conference
S. Kim, B. Feng, K. Smith, S. Masoud, Z. Zheng, C. Szabo, and M. Loper, eds.

HYPERPARAMETER OPTIMIZATION OF DEEP NEURAL NETWORK WITH APPLICATIONS TO MEDICAL DEVICE MANUFACTURING

Gautham Sunder

Supply Chain and Operations Department,
Carlson School of Management,
Minneapolis,USA

ABSTRACT

Bayesian Optimization (BO), a class of Response Surface Optimization (RSO) methods for nonlinear functions, is a commonly adopted strategy for hyperparameter optimization (HPO) of Deep Neural Networks (DNNs). Through a case study at a medical device manufacturer, we empirically illustrate that, in some cases, HPO problems can be well approximated by a second-order polynomial model, and in such cases, classical response surface optimization (C-RSO) methods are demonstrably more efficient than BO. In this study, we propose Compound-RSO, a highly efficient three-stage batch sequential strategy for RSO when there is uncertainty in the complexity of the response surface. Through a simulation study and a case study at a medical device manufacturer we illustrate that Compound-RSO is more efficient than BO for approximating a second-order response surface and has comparable results to BO when the response surface is complex and nonlinear.

1 INTRODUCTION

Due to its state of the art performance deep neural networks (DNNs) are widely adopted for prediction tasks in Industry 4.0, such as, automated quality inspections, predictive maintenance, among others. A decision that is critical to realizing the superlative prediction performance is the choice of hyperparameter settings. For example, prior to training a DNN, the experimenter needs to define model architecture hyperparameters, such as, number of layers and number of nodes in each layer and hyperparameters which control the learning process, such as, learning rate and regularization penalty terms. The problem of searching for the optimal hyperparameter settings that maximize the performance of a DNN is defined as hyperparameter optimization (HPO).

The HPO problem can effectively be modeled as global optimization of an unknown, expensive, and noisy black-box function. Bayesian Optimization (BO), a sequential Response Surface Optimization (RSO) method, is a commonly adopted strategy in the HPO literature. A critical assumption in BO is that the true response surface is complex and that the response surface cannot be well approximated by a second-order polynomial model. In HPO problems, the choice of Matern 5/2 kernel for fitting the Gaussian Process Regression model in BO, reflects the experimenter's prior belief that the hyperparameter response surface is non-smooth and complex (Snoek et al., 2012). However, consistency between the actual complexity of the true response surface and the assumptions made about the response surface complexity is a critical factor in determining the efficiency of the RSO method adopted. For example, when the true response surface is second-order, classical response surface optimization (C-RSO) methods, which are highly-efficient for approximating a second-order response surface, are demonstrably more efficient than BO. However, if the true response surface is complex, C-RSO estimates are biased and, hence, BO is a better suited RSO method. Despite the significance of the validity of the assumptions made regarding response surface complexity, neither C-RSO nor BO formally validate these assumptions.

2 RESPONSE SURFACE OPTIMIZATION

In this study we propose Compound-RSO, a three-stage, batch-sequential RSO strategy for continuous experimental factors, which addresses the above limitation. In the Compound-RSO strategy, we estimate the complexity of the response surface and use the appropriate choice between C-RSO and BO for the estimated response surface complexity. It starts with the second-order assumption and transitions to BO if the second-order approximation is determined to be inadequate. In the first stage of the proposed Compound-RSO strategy, we use Definitive Screening Designs (Jones and Nachtsheim, 2011), which are highly efficient for screening a second-order response surface. In the second stage we augment the initial design with robust runs, which are a compromise between Bayesian D-optimal designs and uniform designs, effectively trading off between D-efficiency and space-filling properties of the design. Our Compound-RSO strategy employs a Lack of Fit (LoF) test to determine the adequacy of the second-order approximation. If significant lack of fit is detected, it implies that the response surface is complex, and hence, we transition to BO in the third stage. Alternatively, if lack of fit is not detected, we propose Adaptive-RSO, an adaptive experimentation strategy for optimizing the second-order response surface in the third stage.

3 RESULTS

We test the efficacy of the proposed Compound-RSO strategy through a simulation study on synthetic test functions of varying complexity. In the simulation study we benchmark the proposed Compound-RSO strategy against BO methods defined in the DiceOptim (Picheny and Ginsbourger, 2014) R package and Spearmint (Snoek et al., 2012) Python library. The results of our simulation study illustrate that the proposed Compound-RSO strategy is demonstrably more efficient than the baseline BO methods when the response surface is second-order and has comparable results to the baseline BO methods when the response surface is complex.

We further evaluate the proposed Compound-RSO strategy through a case study at a large medical device manufacturer on HPO of DNNs used for end of the line quality inspections. As summarized in the Table 1, the DNN model trained at the optimal hyperparameter settings, identified using the proposed Compound-RSO strategy, classifies the four defects of interest with high accuracy.

Table 1: DNN performance at the predicted optimal and poor hyperparameter settings.

Hyperparameter Setting	Loss	Defect1	Defect2	Defect3	Defect4
Optimal Hyperparameter Setting 1	0.1156	99.88%	98.7%	98.34%	94.16%
Optimal Hyperparameter Setting 2	0.1304	99.36%	98.7%	98.88%	92.86%
Poor Hyperparameter Setting	8.0648	76.5%	76.4%	76.9%	77.1%

The contributions of the paper are as follows: (i) we propose robust designs which are supersaturated for a full second-order polynomial model and (ii) we propose a principled approach to estimate the complexity of the response surface and to choose appropriately between C-RSO and BO. Lastly, we illustrate the usefulness of the proposed Compound-RSO strategy through a case study at a medical device manufacturer on HPO of DNNs for quality inspections.

REFERENCES

Jones, B. and Nachtsheim, C. J. (2011). A class of three-level designs for definitive screening in the presence of second-order effects. *Journal of Quality Technology*, 43(1):1–15.

Picheny, V. and Ginsbourger, D. (2014). Noisy kriging-based optimization methods: a unified implementation within the diceoptim package. *Computational Statistics & Data Analysis*, 71:1035–1053.

Snoek, J., Larochelle, H., and Adams, R. P. (2012). Practical bayesian optimization of machine learning algorithms. *Advances in Neural Information Processing Systems*.