

Explainable Hyperparameters Optimization using Hilbert-Schmidt Independence Criterion

PAUL NOVELLO
CEA Cesta - Inria Saclay - CMAP

Supervisor(s): Gaël Poëtte (CEA Cesta), David Lugato (CEA Cesta) and Pietro Congedo (Inria Saclay - CMAP)

Ph.D. expected duration: Nov. 2018 - Nov. 2021

Address: Cesta, 15 Avenue des Sablières, 33114 Le Barp

Email: paul.novello@inria.fr

Abstract:

Computational physics aims at simulating various physical phenomena. These simulations can come with a prohibitive computational cost so that very often they are replaced by reduced models of controlled accuracy [3]. Recently, deep learning has proven its efficiency in several domains, and share similarities with classical reduced models. The idea of replacing some costly parts of simulation codes by an Artificial Neural Network (ANN) thus becomes appealing. To achieve this, the ANN has to reach good accuracy and to remain cost effective (performance-precision trade-off). For both concerns, the hyperparameters (width, depth, activation function, optimizer, learning rate, regularization, etc ...) of the ANN have a strong impact. Hyperparameters have proven to greatly influence accuracy, and some hyperparameters such as their width and depth drive the computational complexity of a neural network.

The number of hyperparameters configurations to test is combinatorial. Besides, most existing approaches for optimizing hyperparameters do not bring any additional comprehension of the individual effect of hyperparameters on the error. They only aim at finding the combination which yields the best error, in a black box fashion. In this work, we consider hyperparameters optimization as a robust system conception task, where the system to optimize is the ANN. A first step is to identify which hyperparameters have the highest influence on the probability of obtaining better ANNs. To that end, we resort to goal oriented sensitivity indices, since we want to assess the importance of hyperparameters towards reaching a certain goal.

Traditional goal oriented sensitivity indices usually require a design of experiments that we can not afford in Hyperparameters Optimization. Indeed, there are very often dozens of hyperparameters to optimize, and a pick and freeze design of experiments would be prohibitive given the training time of ANNs. Instead, we focus on Hilbert Schmidt Independence Criterion (HSIC) [1] and its goal-oriented variant [2]. HSIC is a dependence measure based sensitivity index that can be estimated after a classical Monte Carlo sampling. This drastically limits the computational burden of its estimation. In addition, it can be performed seamlessly after a random search, a classical and popular Hyperparameters Optimization method.

Our contributions are threefold. First, we apply an uncertainty analysis approach, inspired from [2] in the domain of deep learning. Second, we adapt tools from sensitivity analysis (HSIC) to hyperparameters space that can be complex. In particular, hyperparameters can be categorical, continuous, discrete, boolean (activation function vs learning rate), they can interact or may not be always involved in the training (different optimizers parameters), which makes their direct comparison non trivial. We alleviate those concerns. Finally, we describe HO methodologies that make use of the knowledge stemming from this sensitivity analysis and allow satisfying the performance-precision trade-off.

Example In this abstract, we focus on one application to give an idea of the strength of the approach. Let us apply our methodology to the HO of an ANN trained to approximate the solution of the Bateman equations, which lives in R^M . For physical applications, M ranges from tens to thousands. We consider the particular case where $M = 12$. The hyperparameters tested are listed in Figure 1. We sample 1000 different hyperparameters configurations uniformly on the hyperparameters space and estimate HSIC using this Monte Carlo sampling.

For this approximation problem, `learning_rate`, `optimizer`, `activations` and `n_layer` can be considered as highly impactful. Note that HSIC for `n_layers` (depth) is the lowest of the significant HSICs and `n_units` (width) belongs to less impactful hyperparameters. These hyperparameters define the complexity, so the execution speed of the ANN. It means that we do not have to select high values for `n_layers` and `n_units` in order to obtain good errors. To illustrate this remark, we quote that the best of the 1000 initial ANNs (1.99×10^{-4} validation error) has depth 5 and width 470 while another ANN of depth 5 and width 62 reaches 3.74×10^{-4} validation error. This ANN is not so far in terms of error, but much more cost effective since its execution has complexity $\mathcal{O}(n_units^2)$.

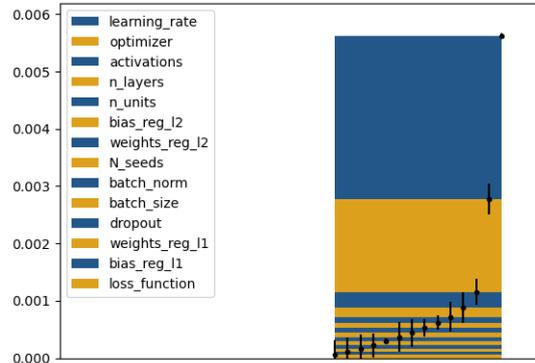


Figure 1: HSIC of hyperparameters for L_2 error

As we saw, this knowledge is valuable by itself. But we can imagine ways of exploiting it further. As examples, we will present a way to select hyperparameters value that improve execution speed without significantly affecting error. We will also introduce a methodology, inspired by [2], which selects most impactful hyperparameters, conduct optimization on the subspace of these hyperparameters, set the values of these hyperparameters at the found optimum and perform a second step of optimization on the remaining hyperparameters. This two step optimization allows simplifying the optimization process by making it focus on the relevant dimensions, but also to select values for non optimized hyperparameters that improve execution speed. Combined to the first application, it allows finding accurate and cost effective ANN. As a highlight we manage to find an ANN of depth 3 and width 32 yielding 3.49×10^{-4} validation error. The number of parameters and FLOPs of the ANN is reduced by a **factor 482** and **380** respectively.

References

- [1] Arthur Gretton, Olivier Bousquet, Alex Smola, and Bernhard Schölkopf. Measuring statistical dependence with hilbert-schmidt norms. In *Proceedings of the 16th International Conference on Algorithmic Learning Theory, ALT'05*, page 63–77, Berlin, Heidelberg, 2005. Springer-Verlag.
- [2] Adrien Spagnol, Rodolphe Le Riche, and Sébastien Da Veiga. Global sensitivity analysis for optimization with variable selection. *SIAM/ASA J. Uncertain. Quantification*, 7:417–443, 2018.
- [3] B. Sudret, S. Marelli, and J. Wiart. Surrogate models for uncertainty quantification: An overview. In *2017 11th European Conference on Antennas and Propagation (EUCAP)*, pages 793–797, March 2017.

Short biography – I graduated from Telecom Paris, with an MSc in Statistics from Imperial College London, in 2018. Then I started a Phd at CEA Cesta and Inria Saclay/CMAP on the application of Deep Learning in a HPC numerical simulation context. More specifically, I study the acceleration of simulation codes with the help of artificial neural networks.