NEURAL ARCHITECTURE SEARCH VIA STANDARD MACHINE LEARNING METHODOLOGIES

GIORGIA FRANCHINI

In the context of deep learning, the more expensive computational phase is the full training of the learning methodology. Indeed, its effectiveness depends on the choice of proper values for the so-called hyperparameters, namely the parameters that are not trained during the learning process, and such a selection typically requires an extensive numerical investigation with the execution of a significant number of experimental trials. The aim of this work is to investigate how to choose the hyperparameters related to both the architecture of a Convolutional Neural Network (CNN), such as the number of filters and the kernel size at each convolutional layer, and the optimisation algorithm employed to train the CNN itself, such as the steplength, the mini-batch size and the potential adoption of variance reduction techniques. The main contribution of this work consists in introducing an automatic Machine Learning technique to set these hyperparameters in such a way that a measure of the CNN performance can be optimised [1]. In particular, given a set of values for the hyperparameters, we propose a low-cost strategy to predict the performance of the corresponding CNN, based on its behavior after only few steps of the training process. To achieve this goal, we generate a dataset whose input samples are provided by a limited number of hyperparameter configurations together with the corresponding CNN measures of performance obtained with only few steps of the CNN training process, while the label of each input sample is the performance corresponding to a complete training of the CNN.

Such dataset is used as training set for a Support Vector Machines for Regression and/or Random Forest techniques to predict the performance of the considered learning methodology, given its performance at the initial iterations of its learning process. Furthermore, by a probabilistic exploration of the hyperparameter space, we are able to find, at a quite low cost, the setting of a CNN hyperparameters which provides the optimal performance. The results of an extensive numerical experimentation, carried out on CNNs, together with the use of our performance predictor with NAS-Bench-101 [2], highlight how the proposed methodology for the hyperparameter setting appears very promising.

References

[1] G. Franchini, V. Ruggiero, F. Porta, L. Zanni, *Neural architecture search via standard machine learning methodologies*, Mathematics in Engineering, pp. 5 (1), 1-21, 2023.

[2] C. Ying, A. Klein, E. Real, E. Christiansen, K. Murphy, F. Hutter, *NAS-Bench-101: Towards reproducible neural architecture search*, Proceedings of the 36–th International Conference on Machine Learning, pp. 7105–7114, 2019.

University of Modena and Reggio Emilia
Email address: giorgia.franchini@unimore.it