
Example: A Classical Algorithm Configuration Approach to Dynamic Learning Rate Configuration

Steven Adriaensen¹ Göktuğ Karakaşlı¹

¹University of Freiburg

Here we solve DAC (Biedenkapp et al., 2020) by reduction to a classical (static) algorithm configuration problem. Specifically, we statically tune the parameters of hand-crafted dynamic learning rate schedulers using SMAC (Hutter et al., 2011). We illustrate this approach for four different schedulers (e.g., see `schedulers.py`):

ConstantLRPolicy: A scheduler using the same learning rate (configurable parameter) throughout the training process.

CosineAnnealingLRPolicy: A cosine annealing scheduler (Loshchilov and Hutter, 2017). Starting from an initial learning rate (configurable parameter) reduces the learning rate to 0 following a re-scaled half cosine curve.

SimpleReactivePolicy: A novel simple scheduler using a constant learning rate (configurable parameter) for the first 2 epochs. After every $(2 + i)^{\text{th}}$ epoch ($i \in \mathbb{N}$), if the average mini-batch loss during the last epoch was lower than that in the previous 2 epochs, it *increases* the learning rate with a factor a (configurable parameter). If the loss is higher, it *decreases* the learning rate with a factor b (configurable parameter).

ReduceLROnPlateauPolicy: A scheduler emulating the ReduceLROnPlateau pytorch scheduler. It adjusts the learning rate per epoch, based on the validation loss (observed at the end of every epoch). In particular, it reduces the learning rate with a ‘factor’ if the validation loss has stagnated for ‘patience’ epochs. This scheduler has many parameters, but we only optimized the initial learning rate, ‘patience’, ‘factor’, and used the pytorch defaults for the others.

We provide the code to configure these schedulers using SMAC (see `train.py`) as well as the resulting optimized configurations (see `tmp/saved_configs`) on 1000 training instances (environment seed was 42) after 5000 target algorithm evaluations (smac seed was 42). Note that the training code requires installing SMAC3 (e.g., calling `pip install -r requirements_train.txt`).

References

- Biedenkapp, A., Bozkurt, H., Eimer, T., Hutter, F., and Lindauer, M. (2020). Dynamic Algorithm Configuration: Foundation of a New Meta-Algorithmic Framework. In Lang, J., Giacomo, G. D., Dilkina, B., and Milano, M., editors, *Proceedings of the Twenty-fourth European Conference on Artificial Intelligence (ECAI’20)*.
- Hutter, F., Hoos, H., and Leyton-Brown, K. (2011). Sequential model-based optimization for general algorithm configuration. In Coello, C., editor, *Proceedings of the Fifth International Conference on Learning and Intelligent Optimization (LION’11)*, volume 6683, pages 507–523.
- Loshchilov, I. and Hutter, F. (2017). Sgdr: Stochastic gradient descent with warm restarts. In *Proceedings of the International Conference on Learning Representations (ICLR’17)*. Published online: iclr.cc.