Countering exploration biases in AlphaZero

ComCo 2021 poster session abstract by Georg Schröter (PhD student at the Institute of Cognitive Science, Universität Osnabrück)

The advent of Deepmind’s AlphaGo approach [1] and its subsequent improvement AlphaZero [2] in 2016 resp. 2018 marks a major milestone in the recent advancements of AI, allowing a neural network agent to be trained in a self-supervised reinforcement learning fashion to master a variety of environments without the intervention of human domain expert knowledge. Despite the general success of the approach both in the original papers and publicly available open source reimplementations [3][4], hyperparameter choices influencing both the necessary RL exploration strategy and the tree search during the creation of training games induce various exploration biases especially in asymmetric situations which require awareness and need to be dealt with before attempting to transfer the AlphaZero approach to inherently asymmetric environments. In our contribution we identify different exploration induced biases displayed by models from the LeelaChessZero project [4] and suggest approaches to reduce or remove said biases in the generation of the training data which would allow the learned neural net model to not be biased by the specific choices of hyperparameters, therefore more closely resembling the behaviour of an optimal agent in the respective environment instead of the often overtly optimistic behaviour typically displayed by Monte Carlo agents.