Searching by learning: Exploring artificial general intelligence on small board games by deep reinforcement learning

Wang, H.

English Summary

In deep reinforcement learning, searching and learning techniques are two important components. They can be used independently and in combination to solve different problems in AI, and have achieved impressive results in game playing and robotics. These results have inspired research into artificial general intelligence (AGI), using these methods. Two general frameworks—General Game Playing (GGP) and AlphaZero—have been built as the testbed to explore different aspects of AGI. Both frameworks combine searching and learning methods.

The purpose of this dissertation is to assess the potential of these methods. We study table based classic Q-learning on the GGP system, showing that classic Q-learning works on GGP, although convergence is slow, and it is computationally expensive to learn complex games. For larger games deep neural networks may work better, which we study next.

Previous work shows that combining searching and learning can achieve better performance. In this approach search is based on the learned neural network model and this model is trained by examples that are played by the search algorithm. This dissertation uses an AlphaZero-like self-play framework to explore AGI on small games. By tuning different hyper-parameters, the role, effects and contributions of searching and learning are studied. In order to understand the relative importance of searching and learning, a correlation experiment is performed, suggesting that within a limited budget, a higher number of the outer self-play iterations is more promising than inner training epochs, search simulations, and game episodes.

A further experiment shows that search techniques can contribute as experts to generate better training examples to speed up the start phase of training. This idea is called *warm-start* in the dissertation. We find that in AlphaZero-like self-play, a combination of Rollout and Rave enhancements can improve the start iterations of self-play training, especially with an adaptive iteration length. The

warm-start method is a promising methods to improve the training by embedding searching techniques in self-play based learning.

Based on the successes of AlphaZero-like self-play in two-player games, we explore the possibility in single-player games. In order to extend the AlphaZero-like self-play approach to single player complex games, the Morpion Solitaire game is implemented by combining Ranked Reward method. Morpion Solitaire is a highly challenging combinatorial puzzle. Our first AlphaZero-based approach is able to achieve a near human best record. This result indicates that AlphaZero-like self-play approach is a promising method to explore AGI in single player games.

Overall, in this thesis, both searching and learning techniques are studied (by themselves and in combination) in GGP and AlphaZero-like self-play systems. We do so for the purpose of making steps towards artificial general intelligence, towards systems that exhibit intelligent behavior in more than one domain. Our results are promising, and propose alternative ways in which search enhancements can be embedded as experts to generate better training examples for the start phase of training.