Position Control and Production of Strategies for Deep Learning Go Programs

1810079 SHI,Yuan

Computer Go programs have surpassed top-level human players by using deep learning and reinforcement learning techniques. Other than the strength, entertaining Go AI and AI coaches are also interesting directions but have not been well investigated. Some researchers have worked on entertaining beginners or intermediate players. One topic is position control, aiming to make strong programs play close games against weak players. Under such a scenario, the naturalness of the moves is likely to influence weaker players' enjoyment. Another topic is producing various strategies (or preferences), which human players usually have. Some methods for the two topics have been proposed and evaluated for a traditional Monte-Carlo tree search (MCTS) program. However, there are some critical differences between traditional MCTS programs and recent programs based on AlphaGo Zero, such as Leela Zero and KataGo. For example, recent programs do not run random simulations to the ends of games in MCTS, making the existing method for producing various strategies not applicable.

In this paper, we first summarize such differences and some resulted problems. We then adapt existing methods as well as propose new methods to solve the problems, where promising results are obtained.

For position control, the modified Leela Zero can play gently against a weaker player (48% of wins against a weaker program, Ray). A human subject experiment shows that the average number of unnatural moves per game is 1.22, while that by a simple method without considering naturalness is 2.29.

We also propose a new position control method specifically for endgames by using expected territory advantages instead of win rates. KataGo, with the proposed new method, can play gently against a weaker player Ray with 37% of wins in endgames while using our previous position control method is 63%.

Finally, for producing various strategies, a new method is introduced. In our experiments, center- and edge/corner-oriented strategies are produced by our method, and human players use five-level evaluation to evaluate the strategies that -2 mean center-oriented strategies and +2 mean edge/corner-oriented strategies. The average evaluation scores of 13x13 center-oriented is -0.5, 13x13 edge/corner-oriented is +0.7, 19x19 center-oriented is -1.25, 19x19 edge/corner-oriented is +0.85. The results show that human players can successfully identify the strategies.