University of Waterloo
CS 486
r2knowle: 2023-11-27

Assignment # 4

Q1)

Q2) For this question we are going to be looking at the paper: "Mastering the game of Go with deep neural networks and tree search."

What are the motivations for the work: As stated in "how to read a research paper", there often exists two aspects of a problem that the paper tries to address. The first aspect is the "people problem" or the anticipated value that a solution could have in the real world. For this paper, this is the benefits artificial intelligence can provide in automation and making complex decisions. Go provides an baseline to demonstrate such an ability. The second aspect of the problem is the "technical problem", which outlines the difficulties and limitations in creating a solution. For Go this is due to the immense search space and difficulty in evaluating the board correctly.

What is the proposed solution: Any game in theory has a perfect method of playing, using optimal values functions we can do an exhaustive search to find it. However with Go, where the search space is so large this would take an impossible amount of time and is not feasible. The proposed solution to reduce the search space is to use value networks to evaluate moves and policy networks to select which moves to use.

What is the evaluation of the proposed solution: In the past solutions utilized a Monte Carol (MCTS) tree search based on shallow policies or a linear combination of input features. Instead this solution passes the board into a CNN as an image and uses the convolutional layers to construct a representation of the position. As a benefit this reduces the effective depth and breadth of the search tree that is necessary. This solution is convincing as citations to performance in image classification, face recognition and Atari games demonstrates he value in this technology.

What are the contributions: The two main contributions are through the value network (used to evaluation positions) and the policy network (used to sample actions). The value network was trained through reinforcement learning, as to approximate the value of a position and thus get as close as possible to the optimal value function (or optimal play). The policy network was trained using policy gradient reinforcement learning, this is done by forcing the current policy to play against previous iterations of the policy network. This also helps to reduce the over fitting issue of running the algorithm on only the training data.

What are future directions for this research: The paper doesn't include a section dedicated to future directions, instead at the end it has a discussion on the results of the research. It mentions how this technology was able to beat the best human go player (5-0) and how it did it in less searches then the previous implementations. It then followed by expressing interest in seeing what other domains AI harnessing this technology can be used in, now that a seemingly impossible task was proven to be solvable.