Artificial Intelligence Nanodegree

Summary of “Mastering the game of Go with deep neural networks and tree search”

Tran Nguyen Bao Trung - 2017.03.16

Paper’s Goal: The paper talks about the new approach to apply deep neural networks and tree search in tackling the game of Go, which is one of the most challenging classic games for AI because of its big search space and the difficulty of evaluating board positions and moves.

AlphaGo’s Techniques: AlphaGo uses “value networks” to evaluate board positions and “policy networks” to select moves. AlphaGo’s neural networks are trained by a hybrid technique that combines Supervised Learning from human expert games and Reinforcement Learning from games of self-play.

“Value networks” are used to predict the outcome from a specific position of games played and trained by Reinforcement Learning. The weights of the value networks are formed by regression on state-outcome pairs, using stochastic gradient descent to minimize the mean squared error (MSE) between the predicted value and the corresponding outcome. However, this naïve approach of predicting game outcomes from data consisting of complete games leads to overfitting. To overcome this problem, new self-play data set consisting of 30 million distinct positions, each sampled from a separate game, is created and each game was played between the Reinforcement Learning policy network and itself until the game terminated.

“Policy networks” are used to predict human expert moves in a data set of positions; they are trained by both Supervised Learning and Reinforcement Learning. A Reinforcement Learning policy networks is initialized to the Supervised Learning policy networks, and is then improved by policy gradient learning to maximize the outcome (winning more games) against previous versions of the policy networks. A new data set is generated by playing games of self-play with the Reinforcement Learning policy network.

Paper’s Result: This is the first time that a computer program has defeated a professional player in the game of Go. The new algorithm that combines Monte Carlo Tree Search with Value and Policy networks achieved a 99.8% winning rate against other Go programs, and defeated the human European Go champion by 5 games to 0.