Research Review

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

Goals

- Develop a Go program to play at the level of the strongest human players
- Introduce a new search algorithm that combines Monte Carlo simulation with value and policy networks

Techniques

- Neural network to reduce depth and breadth of search tree with:
  - Value networks: To evaluate board positions
  - Policy networks: To sample actions

- Deep neural networks trained by
  - Supervised learning by human expert games
  - Reinforcement learning from games of self-play

Neural Network Archiecture

- Neural network to predict human expert moves in a data set of positions:
  - Rollout policy
  - Supervised learning (SL) policy network
    - Alternates between convolutional layers with weights and rectifier nonlinearities
    - Final softmax layer outputs a probability distribution over all legal moves

  - Reinforcement learning (RL) policy network is initialized to the SL policy network
    - Improved by policy gradient learning to maximize outcome
    - New data set is generated by self-play games with RL policy network

  - Reinforcement learning (RL) value network

    - Focused on position evaluation
    - Weights trained by regression to predict expected outcome

    - Searching policy and value network with Monte Carlo tree search (MCTS) algorithm
        - Selects actions by lookahead search

- Policy network
    - Input: representation of board position
    - Convolutional layers with parameters:
        - SL policy network
        - RL policy network
    - Output: probability distribution over legal moves

- Value network
    - Input: representation of board position
    - Convolutional layers
    - Output: expected outcome

Results

- New search algorithm:
    - Achieved 99.8% winning rate against other Go programs
    - Defeated the European Go champion by 5 games to 0