Applying Deep Double-Q Learning and Monte Carlo Tree Search to the Game of Go

Jonathan Booher (jaustinb), Nithin Kannan (nkannan), Enrique De Alba (edealba)
CS221 Final Project, Stanford University

Introduction
Motivation:
• Attempt to improve upon the results of AlphaGo Zero[2] by using Deep Double-Q Learning[1] instead of binary classification of game winner
Problem Definition:
• Replicate the MCTS method used in AlphaGo Zero
• Use Deep Double-Q Learning[1] to predict values and actions

Approaches: Deep Double-Q Learning
Network Design:
• 20 block Residual Network
Deep Double-Q Learning:
• Samples batches of (s,a,r,s') from positions in previous games
• Rewards are +/-1 depending on game winner
• One set of parameters for selecting actions and another for value estimation
  ◦ Only train one set of parameters at a time
  ◦ Swap the parameters every 3 training iterations
• Formulas
  \[\theta_{t+1} = \theta_t + \alpha(Y_t^Q - Q(S_t,A_t;\theta_t)) \nabla_{\theta_t}Q\]
  \[Y_t^Q \equiv R_{t+1} + \gamma \max_{a'} Q(S_{t+1}, a'; \theta_t')\]

Approaches: Monte Carlo Tree Search
Monte Carlo Tree Search (MCTS):
• Used to generate games of self play from network parameters
• PUTC used for exploration
• Noise at root for more exploration
Self Play:
• Moves based on visit count in MCTS
• Temperature annealing ensures more exploration in the opening

Challenges
Computational:
• High Branching Factor
• Computing Gradients on large network
  ▷ Play on 9x9 Go board
  ▷ Tensorflow across 2 Titan Xp GPUs
Algorithmic:
• Setting hyper-parameters
• Initialization of parameters (overflow)
  ▷ Decreased computation time allows quicker parameter optimization

Results
Testing:
• Tested by playing against the Go engine Pachi using a modified OpenAI Gym[3]
  • ELO difference calculated based on our win rate. Our ELO is currently 2280.
  ![ELO Difference vs Epoch](image)

Analysis
Convergence:
• Loss appears to converge
Performance:
• Limited by quality of data
• Improved by deeper MCTS and more simulations
  ![Loss vs Epoch](image)

References