Gelei Chen
March 4, 2017

AlphaGo Research Review

Article: Mastering the game of Go with deep neural networks and tree search
Link: https://storage.googleapis.com/deepmind-media/alphago/AlphaGoNaturePaper.pdf

Goal: Build a computer Go agent that can beat best human player using both 'value networks' and 'policy networks'

Abstraction: The game of Go is used to be considered impossible for artificial intelligence because of its massive search space and the trouble of judging board positions as well as moves. DeepMind team uses 'value networks' as an advanced heuristic function, and 'policy networks' as an advanced version of alpha-beta pruning or mini-max algorithm. DeepMind team uses both supervised learning from human expert games and reinforcement learning from games of self-play.

Key words:

- Value networks:
  - The purpose of 'value networks' is same as our heuristic function. DeepMind says that 'Ideally, we would like to know the optimal value function under prefect play. In practice, we instead estimate the value function for our strongest policy, using the RL policy network'. I think it is smart that they approach this evaluation problem in an indirect way.
- Policy networks:
  - Supervised learning of policy networks(SL) and Reinforcement learning of policy networks(RL) are the first and second stage of their training pipeline. SL is learned from human Go expert. RL is learned by self-playing. Deep neural network is improved through those two steps instead of our hard-coded alpha-beta pruning or mini-max algorithm.
- Monte Carlo algorithms:
  - Use random algorithm and statistics to give reasonable estimates about some hard to answer questions. The tricky part is the prior as well as the posterior probabilities. In the case of Go, since the winning probability at each position is so small, DeepMind team may need lots of random samplings and customized smoothing techniques, which are limited by infrastructure like network bandwidth and computer hardware. A classical Monte Carlo algorithm consists a big loop that includes selection, expansion, evaluation and backup.
- Deep neural networks:
  - Interconnected web of notes, and edges that join them together. Receive a set of inputs, perform complex calculation in the middle layer or hidden layer and then use output to solve a problem. For classification problem, I can use each layer to extract different features that interests me. (The layer inside can be non-linear.)

- Supervised learning:
    - Human Go expert plays against DeepMind AlphaGo agent. And interact with the system by telling the goodness of each game. This SL policy network can give acceptable response in very short period of time.
- Reinforcement learning:
    - AlphaGo plays against itself for self-learning and self-improvement. This RL policy network improves the SL policy network by optimizing the final outcome of games of self-play.

Conclusion

Google DeepMind has made exciting progress in the Go world. The ways that they structure the solution, take advantage of deep neural network, gather as well as formatting and cleaning training data, optimizing agent performance in a hardware level and training AlphaGo agent in such short period of time, show that DeepMind team is really at generalizing this AI technique and is ready to apply it to other fields.