1、Select a Game-Playing paper：

AlphaGo by the DeepMind Team.

2、Write a simple one page summary of the paper covering the following:

➢ A brief summary of the paper's goals or techniques introduced (if any).

Owing to its enormous search space and the difficulty of evaluating board positions and moves. The game of Go is the most challenging of classic games for artificial intelligence. Paper's goals are to reach the top level of the game of Go and achieve a highest winning rate. Surprisingly, they did not mention any professional Go knowledge and introduce several simple algorithmic techniques about how to solve the problems:

a、 A new approach of two algorithm framework, 'value networks' to evaluate board positions and 'policy networks' to select moves. In search algorithm they combine Monte Carlo simulation with value and policy networks, and develop the program base on a combination of deep networks and tree search.

b、 Employ CNN architecture for the game of Go to construct increasingly abstract, localized representations. They use convolutional layers to construct a representation of the position with a 19*19 image about the board position, and reduce the effective depth and breadth of the search tree.

c、 In the first stage of policy networks, they use supervised learning to predict expert moves from human's data, and trained policy network from the KGS Go Server from 30 million positions. Larger networks achieve better accuracy but are slower to evaluate during search.

In the second stage, they improve the policy network by reinforcement learning. They play games between the current policy network and a randomly selected previous iteration of the policy network from a pool of opponents, it's a good way stabilizes training by preventing overfitting. In the evaluation, RL policy networks achieve a high winning rate when contrasts other strategies.

d、 The value networks has a similar architecture to the policy network , and use Reinforcement learning. There are some approaches to prevent overfitting when constructing training data. When trained on the KGS data of complete games the value network memorized the game outcomes rather than generalizing to new positions and leads to overfitting. To mitigate this problem, generated a new self-play data set consisting of 30 million distinct positions and each sampled from a separate game that RL policy network self-play until terminated.

- A brief summary of the paper's results

  In the evaluation of the level of the game of Go, AlphaGo achieved a 99.8% winning rate against other Go programs, and defeated the human European Go champion, a feat previously thought to be at least a decade away.