Playing Startcraft II with Reinforcement Learning

Team member(s): Haiyuan Mei (hmei0411@stanford.edu)

Motivation

Blizzard's Startcraft series game is one of the most challenging Real Time Strategy game, and has gained immense commercial and cultural success over the past 20 years. However, unlike its popularity, the study of AI in solving full SC games progresses somewhat slow and seems relatively far from finished, this is probably because of the challenge imposed by multi-agent interacting, imperfect information, large state/action space and delayed credit assignment requiring long-term strategies over thousands of steps. The most recent news is about Deepmind's AlphaStar in Jan/2019, which was defeated by Pro SCII champ Grzegorz "MaNa".

As a big fun of Starcraft game myself I would like to take this opportunity to start myself in this exciting journey of solving SC2 game. The study of SC2 also opens a door to unparalleled opportunities to many challenging frontiers, both in AI research and in real life. As a starting point of my study of SC2, this course project is to some extent an attempts to replicate DeepMind's paper of SC2LE (StarCraft II: A New Challenge for Reinforcement Learning) [1].

The scope of the project

As MDP is used to model the problem, with states/actions defined below:

- States: An abstracted RGB images seen during human play which is called "feature layers", which maintains the core spatial and graphical concepts of SC2 (screen and minimap); plus various non-spatial information such as gas and minerals collected, set of actions currently available, etc.
- Actions: A close mimic of human interface with keyboard and mouse, defined as a function identifier plus 0 or more parameters. Such as selecting multiple units can be just denoted as '[select_rect, [[select_add], [x1,y1],[x2,y2]]]'
- Rewards: Two rewards described in the above paper, ternary 1 (win) / 0 (tie) / -1 (loss) received at the end of a game (with all-zero rewards during the game), and Blizzard score. In this course project only the Blizzard score will be explored.

[1] StarCraft II: A New Challenge for Reinforcement Learning **https://arxiv.org/abs/1708.04782**

[2] Asynchronous Methods for Deep Reinforcement Learning https://arxiv.org/abs/1602.01783

In order to evaluate the project, the baseline will be played by a novice SC2 player and should be defeated by my agent; I myself as an intermediate SC2 player will be the oracle. I will initially pick 2 of the mini games in the project:

- MoveToBeacon: The agent has a single marine that gets +1 each time it reaches a beacon.
- FindAndDefeatZerglings: The agent starts with 3 marines and must explore a map to find and defeat individual Zerglings. This requires moving the camera and efficient exploration.

In order to work with video frames, CNN is an important part in this project. The reinforcement learning algorithm will be Asynchronous Advantage Actor Critic (A3C) as described by Mnih et al.[2] Based on how much time left I would also be very interested in applying q-learning and TD learning in this part.

[1] StarCraft II: A New Challenge for Reinforcement Learning **https://arxiv.org/abs/1708.04782**
[2] Asynchronous Methods for Deep Reinforcement Learning https://arxiv.org/abs/1602.01783