Oshan Karki 05/02/2019

CS475 (Machine Learning)

Robert Baker Heckendorn

Report on: **Creating Pro-Level AI for Real-Time Fighting Game with Deep Reinforcement Learning**

Reinforcement Learning have surpassed human level performance in a player vs Environment setting using Value based and Actor-Critic methods. However, for a real-time fighting game in a player vs player setting where each player makes their skill decision simultaneously in real-time, an RL agent should be able to deal with the opponent's skill decision that is unknown at decision making time. Researchers are looking for possible solution to this problem which can make the agent react accordingly to any of the player's strategies in real-time.

To overcome such problem, researchers invented a self-learning algorithm which uses self-play curriculum which introduces opponents with different fighting styles. Fighting styles are made using technique called reward shaping where a supplemental reward is provided to make a problem easier to learn which is used to create three different fighting styles: Aggressive, Balanced and Defensive. The algorithm makes these three-fighting styles reinforce together which helps agents compete against various opponents with real-time decision-making ability. The RL learning process boosts good actions and suppress bad actions so that every time agents makes better action over players. Researchers used ACER algorithm which enables them to deal with policy lag between simulator and learner. They made use of three learning processes each consisting of a learner and 100 simulators such that if an agent is trained for a week, it is equivalent to two years of gameplay experience. The learning process modifies the policy such that good action is prioritized over bad action. This action is only possible with Self-Play curriculum which uses opponent pool for training. In the initial stage of Self-Play curriculum, opponents and agents reinforce together which helps improve agent's training experience and opponent's generalization performance. Agent performs well against single opponent with fixed fighting pattern but fails to deal effectively against other diverse opponents. Algorithm also makes use data skipping technique such as no-op and Maintaining move action. No-op technique will automatically skip overused skill and do nothing for certain period until the skill is available again. Maintaining move action will allow agent to make literal moving decision and lead to change the subsequent states of movement which is useful because the chance of random policy making the same decision consecutively is very low.

The researchers solved the problem by providing pool of opponent data set which defines opponent's nature: aggressive, balanced or defensive using reward shaping. Later, they used those data set to train agents by making them compete against each other. Through the shared pool of learning data every agent(made up of multiple neural nets) encounters different types of opponent and over time iteratively learns to deal with them. I believe they were successful because agents trained against different nature of opponents gave them 62%-win rate against professional gamers. The one issue they had is, they were unable to evaluate AI-performance when in aggressive mode, this could be because human players often need breaks between fights to plan their move strategically, this could be the possible weakness in their whole experiment. Researchers also had to deal with continuously changing skill set in a fight, which they solved using no-op data skipping technique.