

Multiagent reinforcement learning in the Iterated Prisoner's Dilemma

Tuomas W. Sandholm^{*1}, Robert H. Crites²

University of Massachusetts at Amherst, Computer Science Department, Amherst, MA 01003, USA

Reviewer: Gyuseok Lee

Date: 09/26/2022

- Overview of paper content (0.5 points)

This paper assumes that Iterated Prisoner's Dilemma, where the agents' payoffs are neither totally positively nor totally negatively correlated. Plus, it investigates the ability of a variety of Q-learning agents to play the IPD game against an unknown opponent. As a result, all the learners faced difficulties when playing against other learners, agents with longer history windows, lookup table memories, and longer exploration schedules fared best in the IPD games.

- Strengths (0.5 points)

This paper addressed the IPD game problems by studying Q-learning. As a result, it proves that agents with longer history windows, lookup table memories, and longer exploration schedules fared best in the IPD games.

- Weaknesses (0.5 points)

I think that lookup table memories would be exponentially increasing when the number of agents and actions increases. So it would be critical problem for applying the method to real word problem.

- Questions or Discussion (0.5 points)

I wonder what if transformer used for capturing the history instead of lookup table or recurrent unit s. Would it get better result?