Convergence and Performance of Multi-Agent Q-Learning

Supervisor(s):
Francesco Belardinelli (Department of Computing, francesco.belardinelli@imperial.ac.uk)

Project description:
Multi-agent learning algorithms have been shown to display complex, unstable behaviours in a wide array of games. In fact, previous works indicate that such non-stationary behaviours are commonplace and are more likely to occur as the total number of agents increases. This seemingly prohibits convergence to stable strategies, such as Nash Equilibria, beyond special game classes such as zero-sum games or potential games. This project builds upon recent advances in this area [1-6] to analyse the Q-Learning dynamics, a classical model which describes the behaviour of agents who explore their space of actions whilst simultaneously aiming to maximise their rewards. Our goal is to describe the behaviour of Q-Learning in stateful game structures, including Markov Decision Processes (MDPs) and Markov games, and to understand the factors which contribute towards its convergence.

Goals:
1) The first goal is to show that Q-Learning can converge to a unique stable strategy, namely the Quantal Response Equilibrium (QRE), in arbitrary Markov games.
2) We will study Markov games where interactions between agents are constrained by a network to determine under which conditions, depending on the game and network structure, the agents converge to a unique QRE. We will compare the QRE with the underlying Nash to check whether it is an ε-approximation of the Nash Equilibrium.
3) We will perform a statistical analysis of the dynamics of Q-Learning in network Markov games. We will parametrise network Markov games in terms of correlations between agent payoffs and study the average behaviour of the Q-Learning dynamics across all games drawn from a choice of this parameter, thus establishing choices of parameters for which Q-Learning dynamics converge to a stable fixed point.
4) We will study the non-stationary behaviours of Q-Learning, with the aim to provide guarantees which do not rely on asymptotic convergence. We aim to show that Q-Learning dynamics converges to within a neighbourhood of a QRE in network games.
5) We will analyse the performance of Q-Learning dynamics. Here, we will consider a number of cases which suggest that Q-Learning agents may achieve a higher Social Welfare by following non-convergent trajectories, than by playing a stable strategy. We will run experiments to corroborate the theoretical findings.

Timeline (tentative):
Jan 2025: literature review/preliminary experiments as well as completion of task (1)
April 2025: completion of task (2)
June 2025: completion of tasks (3) and (4).
July 2025: tackling the stretch goal (5), write up of final report.

Minimum viable thesis:
The references [1]-[6] have already proved preliminary results for stateless network games. A minimum viable thesis would apply the same experimental methods to stateful MDPs and Markov games and possibly extend the most basic results on convergence.

Required background & skills:
Game Theory, Reinforcement Learning, Control Theory, a solid mathematical background.

Representative References:

[1] A. Hussain, F. Belardinelli, and G. Piliouras, “Asymptotic convergence and performance of multi-agent q-learning dynamics,” pp. 1578–1586, International Foundation for Autonomous Agents and Multia-gent Systems, 2023.
[2] A. Hussain, D. Leonte, F. Belardinelli, and G. Piliouras, “On the stability of learning in games with many players [under review],” pp. 1578–1586, International Foundation for Autonomous Agents and Multiagent Systems, 2024.
[3] A. Hussain, D. Leonte, F. Belardinelli, and G. Piliouras, “Stability of multi-agent learning: Convergence in network games with many players,” in ICML Workshop on New Frontiers in Learning, Control, and Dynamical Systems, 2023.
[4] A. Hussain and F. Belardinelli, “Stability of multi-agent learning in competitive networks [under review],” in Proceedings of the 2024 AAAI/ACM Conference on AI, Ethics, and Society, (New York, NY, USA), Association for Computing Machinery, 2024.
[5] A. Hussain, F. Belardinelli, and D. Paccagnan, “The impact of exploration on convergence and performance of multi-agent q-learning dynamics,” vol. 202, pp. 14178–14202, PMLR, 7 2023.
[6] A. A. Hussain, F. Belardinelli, and G. Piliouras, “Beyond strict competition: Approximate convergence of multi-agent q-learning dynamics,” in Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI 2023, 19th-25th August 2023, Macao, SAR, China, pp. 135–143, ijcai.org, 2023.