Paper Review: Multi-Agent Reinforcement Learning: A Selective Overview of Theories and Algorithms

Tomoki Koike

***Abstract*—We will analyze and dissect a paper by Kaiqing Zhang,Zhuoran Yang, and Tamer Bacsar [1]. This paper has its emphasis on defining the theoretical foundations, frameworks, and algorithms of Reinforcement Learning convolving with multi-agent systems. This is known as multi-agent RL (MARL) and this paper will review the Markov/stochastic and extensive-form structure games that are played by multiple agents. By providing the basis of theories and algorithms for MARL, this paper aims to allow the fruition of other research to topics and advances in the application and challenges of MARL.**

***Index Terms*—Reinforcement Learning (RL); Multi-agent RL (MARL); Markov Chain; Stochastic; Q-learning; dynamic programming; game theory; decentralized control**

I. Introduction and Motivation

This paper has adopted its idea of research from the recently rise in development of reinforcement learning and neural networks (DNNs) that play a board game, drive autonomous vehicles, and conduct measurements with sensor communication. Most of the applications happen to be possible to model as a single independent system, however, the game and application itself is truly a multi-agent system. Thus, the paper delved into the idea of constructing a multi-agent reinforcement learning algorithm to accelerate the development of multi-agent problems which involve cooperativeness or competitiveness between agents.

The paper explains how there are three large groups in MARL: fully cooperative, fully competitive, and a mix of those two. In a cooperative setting the agents collaborate to optimize a common long-term return or in other words work reciprocally to achieve a convergence or consensus. Whereas in a competitive setting the agents status add up to become zero at the end. Since the objective are not mostly aligned it is a challenge to develop a learning algorithm to reach an equilibrium. Though a single agent RL is also not fully scrutinized the paper provide a roadmap to MARL and its theories and algorithms.

Purdue University AAE

II. Problem Formulation

To understand MARL, it is essential to first understand single-agent RL. RL performs a sequence of decision-making through the interaction of the environment and processing the progress with rewards. The decision-making is formulated as a Markov decision process (MDP) consisting of a state-action function/Q-function, reward function, and updating process with system transition probability. With RL of single-agents there are two mainstream types which are value-based and policy-based methods. Value-based methods find a best estimate of the state-action value function which is also known as the Q-function to optimize the performance. The policy-based method uses parameterized function approximations like neural networks which is used to search within the policy space which allows the algorithm to find the gradient or fastest path to the long-term reward which is then updated along that direction.

Once the paper has elucidated the background of single-agent RL, the discussion moves onto the MARL frameworks. Namely Markov/stochastic games and extensive-form games. The former system has each agent work independently having different rewards and action but the action is chosen simultaneously, and this system can only handle fully observed cases. The latter can handle imperfect information models where the environment is not fully understood by having a special agent determine the randomness of the environment adhering to a policy and the other agents follow a sequence of actions that contain the history of actions and allow a coherent action by agents to some extent. For each method there are more preferable settings either being cooperative or competitive and there are a variation of models and theories formulated for each specific case.

III. Main Results

After covering the theories behind RL and MARL, the paper begins to posit the challenges involving MARL. First challenge is the unclarity or ambiguity of the goals. This is an obstacle of the learning process and it is even questionable that convergence is the main goal of MARL algorithms. It is hypothesized that it is rather important to consider a cyclic convergence and separating the learning goal into a criteria of being stable and rational. The second challenge is that controlling a multiple learning processes concurrently acts counterproductive when it is required to have one agent behave and adapt accordingly to other agents in the learning process and not independently for a MARL. The third one is the scalability issue that rises when the there are multiple agents and the computation might be too intensive with a significantly high dimensions of analysis. Finally, the cooperative nature of multi-agents complicate the information structure based on numerous amount of edges that connect the communication network of the agents.

IV. Your Ideas of Further Improvements

This reading was quite overwhelming in its number of words as well as the content. The fourth chapter including all the proposed algorithms for each framework, method, and settings were broadly covered and was convoluting. However, this paper did provide me with the foundations of MARL and showed me where to get started. The successful applications in autonomous aerial vehicles and game playing were very interesting and left many potential of further investigation.

References

[1] Kaiqing Zhang, Zhuoran Yang, and Tamer Ba**c**sar. Multi-agent reinforcement learning: A selective overview of theories and algorithms. *Handbook on RL and Control (Springer Studies in Systems, Decision and Control)*, 1, 2019.