2024

Literature Review

by Dmitry Beresnev, d.beresnev@innopolis.university
and Vsevolod Klyushev, v.klyushev@innopolis.university

Selected papers

- Zhang, K., Yang, Z., & Başar, T. (2021). Multi-agent reinforcement learning: A selective overview of theories and algorithms. Handbook of reinforcement learning and control.
- Foerster, J., Assael, I. A., De Freitas, N., & Whiteson, S. (2016). Learning to communicate with deep multi-agent reinforcement learning.
- ~~Gronauer, S., & Diepold, K. (2022). Multi-agent deep reinforcement learning: a survey.~~
- Canese, L., Cardarilli, G. C., Di Nunzio, L., Fazzolari, R., Giardino, D., Re, M., & Spanò, S. (2021). Multi-agent reinforcement learning: A review of challenges and applications.
- Li, S., Wu, Y., Cui, X., Dong, H., Fang, F., & Russell, S. (2019, July). Robust multi-agent reinforcement learning via minimax deep deterministic policy gradient.
- M. Wen et al., (2022). Multi-Agent Reinforcement Learning is a Sequence Modeling Problem.

Multi-agent reinforcement learning: A selective overview of theories and algorithms

Introduction

The paper provides a comprehensive review of Multi-Agent Reinforcement Learning (MARL), which deals with reinforcement learning (RL) in environments with multiple agents. MARL addresses complex multi-agent settings where agents interact, cooperate, or compete in shared environments. Though empirically successful, theoretical foundations for MARL are relatively lacking in the literature. The authors aim to offer an overview of MARL algorithms grounded in theoretical foundations, specifically within frameworks like Markov/stochastic games and extensive-form games. Moreover, authors provide several significant applications of MARL (Unmanned Aerial Vehicles, Texas Hold'em Poker, etc.), and supplement the MARL taxonomy with several new angles.

Multi-agent reinforcement learning: A selective overview of theories and algorithms

Methodology

The paper highlights two key frameworks for MARL: Markov games and extensive-form games. **Markov games** generalize Markov Decision Processes (MDPs) to the multi-agent setting, incorporating scenarios where agents interact and influence rewards of each other and the environment dynamics. **Extensive-form games** are used when agents have imperfect information about the environment. The paper also categorizes MARL tasks into fully cooperative, fully competitive (zero-sum game), and mixed settings (general-sum game), each requiring different algorithmic approaches and theoretical considerations. The authors review existing algorithms, especially emphasizing ones with theoretical guarantees, including Q-learning, policy-gradient methods, and decentralized approaches. They introduce taxonomies underexplored in existing reviews, such as decentralized MARL and the mean-field regime.

Multi-agent reinforcement learning: A selective overview of theories and algorithms

Validations and results

The authors evaluate the performance of MARL algorithms by considering their theoretical **convergence properties**, **complexity**, and **applicability** to various tasks. Paper discusses the convergence of algorithms in terms of finding Nash equilibria (the point that no rational agent will deviate from, if algorithm finally converges) and finding best learning strategy for a given agent and a fixed class of the other agents in the game. The paper emphasizes policy-based methods (such as Actor-Critic) for their better convergence properties compared to value-based methods, especially when using function approximation, e.g. neural networks. Additionally, the authors explore decentralized MARL, highlighting its relevance for real-world applications like sensor networks and intelligent transportation systems, where agents learn in a distributed manner without centralized control.

Multi-agent reinforcement learning: A selective overview of theories and algorithms

Strengths

The main strength is structured review of MARL algorithms, specifically focusing on methods supported by theoretical analysis. Paper provides valuable insights into the challenges of MARL, such as non-stationarity, scalability, and the complexity of decentralized learning. Authors review both well-established algorithms and newly emerging directions.

Weaknesses

However, while focusing on algorithms with theoretical guarantees, paper leaves out a broader spectrum of empirically successful but theoretically unproven methods, such as deep MARL algorithms. The authors proposed the the theoretical analysis of deep MARL to the future researches.

Learning to Communicate with Deep Multi-Agent Reinforcement Learning

Introduction

The authors address a key challenge in MARL: how autonomous agents can develop communication protocols to coordinate their actions effectively in cooperative tasks. Traditional MARL often assumes predefined communication protocols, but considered paper focuses on the learning of communication strategies among agents. The central problem is enabling agents to learn how to share information over limited-bandwidth channels in partially observable and noisy environments, where no agent has a complete view of the state, and communication becomes necessary for achieving a shared goal. Within this research, authors are focused on settings with centralised learning but decentralised execution: munication between agents is not restricted during learning, but during execution of the learned policies, the agents can communicate only via the limited-bandwidth channel.

Learning to Communicate with Deep Multi-Agent Reinforcement Learning

Methodology

The authors propose two learning approaches for MARL communication tasks: Reinforced Inter-Agent Learning (RIAL) and Differentiable Inter-Agent Learning (DIAL). **RIAL** utilizes deep Q-learning to allow agents to learn individual Q-values for both environment actions and communication actions. It enables agents to communicate indirectly by learning communication protocols through their interactions with the environment. However, in RIAL agents do not give each other feedback about their communication actions . **DIAL** liquidates this limitation by introducing a protocol where agents communicate via continuous signals during training, enabling gradient backpropagation across agents. Hence, DIAL directly optimizes communication channels.

Learning to Communicate with Deep Multi-Agent Reinforcement Learning

Validations and results

The validation of the proposed methods isdone through experiments in two benchmark settings **Switch Riddle** and **MNIST games**.

In the Switch Riddle, a classic multi-agent task requiring coordination, DIAL showed faster convergence to an optimal policy.

In the MNIST games (which are, basically, two games: Colour-Digit and Multi-Step MNIST) agents had to communicate encoded information to classify images.

Their experimental results show that DIAL, especially with parameter sharing, significantly outperforms RIAL and non-communication baselines in learning communication protocol.

The authors also evaluated the impact of noise in communication channels and demonstrated that adding noise during training helps agents to maintain discrete and effective communication protocol.

Learning to Communicate with Deep Multi-Agent Reinforcement Learning

Strengths

The main strength is development of RIAl and DIAL architectures. Moreover, DIAL is is the first approach to implement differentiable communication across agents in deep MARL. Also, the paper introduces two well-described environments that are well-suited for studying communication in MARL. Finally, the research objective of the paper, communication through limited-bandwidth channels, is particularly relevant for real-world applications, such as robotics and network settings.

Weaknesses

However, paper focuses only on small, synthetic environments, what gives almost no insights about performance of proposed methods in more complex, real-world settings. Despite success of DIAL, it relies on centralized learning during training, which can be crucial limitation in practical tasks. Also, as DIAL uses gradient-based optimization through continuous communication, it may not generalize to tasks requiring fully discrete communication.

Multi-agent reinforcement learning: A review of challenges and applications

Introduction

The paper provides an overview of Multi-Agent Reinforcement Learning (MARL), focusing on its challenges and applications. It highlights the growing importance of MARL in real-world tasks where multiple agents interact with their environment, such as traffic management, robotics, and telecommunications, stressing the need to address issues like non-stationarity, scalability and observability in multi-agent systems.

Non-stationary environment - as agents act simultaneously, each agent's action can alter the environment, complicating the learning process for others.

Scalability - handling large numbers of agents

Observability - how much of the environment an agent can perceive.

The authors aim to provide an organized review of existing MARL algorithms, their structures, and their applications

Multi-agent reinforcement learning: A review of challenges and applications

Methodology

The authors reviewed key MARL algorithms, dividing them into categories based on their structure and approach. They described several frameworks, including **Markov Decision Processes** and **Markov Games**, which model environments for single and multi-agent learning. The paper analyzes both **value-based** and **policy-based** MARL algorithms, including **actor-critic** approach.

The methodology also includes models for dealing with partial observability, where agents cannot perceive the entire environment. Different approaches like **Partially Observable Markov Decision Processes (POMDPs)** and **Decentralized POMDPs (Dec-POMDPs)** are explored to handle uncertainty in the environment.

Multi-agent reinforcement learning: A review of challenges and applications

Validations and results

The paper validates the performance of the discussed MARL algorithms by applying them to various benchmark environments that simulate real-world scenarios, such as traffic control systems, robotic control systems, and complex video game environments like **StarCraft II**.

It emphasizes challenges like **non-stationarity** and **partial observability**, and discusses performance in terms of **scalability** and **applicability** to real-world problems.

For instance, Q-learning-based algorithms, despite being popular, often struggle with convergence in multi-agent settings. (When Deep Q Networks and Asynchronous Advantage Actor-Critic show better adaptability).

Multi-agent reinforcement learning: A review of challenges and applications

Strengths

The paper offers a comprehensive taxonomy of MARL algorithms and explores their performance across **various benchmark** environments. Paper provides clarity on how different algorithms handle **multi-agent complexities** such as **non-stationarity** and **scalability**.

Weaknesses

A noted limitation is the lack of **theoretical guarantees** for some MARL approaches, especially in non-stationary and partially observable environments. Additionally, many algorithms struggle with **scalability** as the number of agents increases. The **high computational cost** and the need for **vast amounts of data** for training deep learning models in MARL are other noted challenges.

Robust multi-agent reinforcement learning via minimax deep deterministic policy gradient

Introduction

The paper focuses on developing a robust multi-agent reinforcement learning (MARL) algorithm. It highlights how traditional deep reinforcement learning (DRL) methods, while successful in single-agent scenarios, struggle in multi-agent settings due to training instability. As agents co-evolve in a non-stationary environment, their policies can get stuck in suboptimal local solutions. The authors propose a new approach, **Minimax Multi-Agent Deep Deterministic Policy Gradient (M3DDPG)**, aimed at creating more robust policies that generalize well when facing opponents with different strategies.

Robust multi-agent reinforcement learning via minimax deep deterministic policy gradient

Methodology

The core of the paper's methodology involves the development of **M3DDPG**, an extension of the **Multi-Agent Deep Deterministic Policy Gradient (MADDPG)** algorithm.

The **M3DDPG** introduces a **minimax approach** for robust learning, where agents optimize their policies by considering the worst-case responses from opponents. **The Multi-Agent Adversarial Learning (MAAL)** framework is introduced to make the minimax learning objective computationally feasible by approximating it with a **single** gradient descent step.

Robust multi-agent reinforcement learning via minimax deep deterministic policy gradient

Validations and results

The **M3DDPG** algorithm is tested on four different mixed cooperative and competitive environments, such as **Covert Communication**, **Keep-away**, **Physical Deception**, and **Predator-Prey**. The results show that agents trained using **M3DDPG** significantly outperform those trained using **MADDPG** in terms of robustness. In competitive settings, **M3DDPG** agents adapted better to adversarial opponents and maintained stronger performance even under worst-case scenarios.

Robust multi-agent reinforcement learning via minimax deep deterministic policy gradient

Strengths

One of the key strengths of the paper is its introduction of the **minimax** optimization technique in MARL, which enhances the robustness of the agents' policies. The **MAAL** framework also allows for efficient computation, enabling practical implementation of the otherwise computationally expensive minimax problem. Additionally, the comprehensive empirical evaluations across various environments demonstrate the effectiveness of the proposed algorithm.

Weaknesses

The paper acknowledges that despite the robustness provided by M3DDPG, the one-step gradient approximation in MAAL limits the exploration to only locally worst situations. This can lead to **suboptimal** behavior when agents face opponents with drastically different strategies. The algorithm may require further refinement to handle broader adversarial scenarios and to better balance robustness with computational efficiency.

Multi-Agent Reinforcement Learning is a Sequence Modeling Problem

Introduction

The paper introduces the **Multi-Agent Transformer** (MAT) model, which redefines Multi-Agent Reinforcement Learning (MARL) as a sequence modeling problem. The authors argue that MARL can benefit from the advancements in sequence models (SMs), such as the **Transformer** used in **natural language processing** (NLP), by viewing multi-agent decision-making as a sequential task. They introduce MAT, an architecture that uses a **sequential encoder-decoder** framework to map agent observation sequences to optimal action sequences, leveraging advantage decomposition to transform the policy search problem into a sequential decision-making process.

Multi-Agent Reinforcement Learning is a Sequence Modeling Problem

Methodology

The authors propose the **Multi-Agent Transformer** (MAT), which combines a Transformer-based architecture with the multi-agent advantage decomposition theorem. The MAT architecture includes an **encoder** that processes a sequence of agent observations and a **decoder** that sequentially generates **optimal actions** for each agent. The design allows for efficient **on-policy learning** through **self-attention** mechanisms while reducing the computational complexity of MARL problems to a linear scale. MAT is trained online, and its design supports joint policy optimization without requiring pre-collected data.

Multi-Agent Reinforcement Learning is a Sequence Modeling Problem

Validations and results

The MAT model was tested on several benchmark environments, including **StarCraft II**, **Multi-Agent MuJoCo**, and **Google Research Football**. The results demonstrated that MAT **outperforms** strong baselines such as MAPPO, HAPPO, and QMIX in terms of data efficiency and performance across both **cooperative** and **competitive** tasks. Additionally, MAT showed excellent few-shot learning capabilities, adapting well to **unseen** tasks with minimal data.

Multi-Agent Reinforcement Learning is a Sequence Modeling Problem

Strengths

MAT offers several strengths:

- Effectively models multi-agent systems by leveraging the Transformer architecture.
- Architecture allows for parallel training while ensuring monotonic performance improvement, providing high sample efficiency.
- The model's perform well on few-shot and unseen tasks, making it versatile in different environments.

Weaknesses

One key limitation is the **complexity of scaling** MAT to very large environments (with a high number of agents). Additionally, the performance of MAT might vary based on the specific tuning of **hyperparameters**, which could limit its applicability across a broad range of tasks without optimization.

Thanks!

Do you have any questions?