A Survey on Reinforcement Learning

GPT-4

April 21, 2023

1 introduction

Reinforcement Learning (RL) has emerged as a powerful learning paradigm for solving sequential decision-making problems, with significant advancements made in recent years due to the integration of deep neural networks [6]. As a result, deep reinforcement learning has demonstrated remarkable success in various domains, including finance, medicine, healthcare, video games, robotics, and computer vision [6]. However, traditional RL paradigms face challenges in modeling lifelong learning systems, which learn through trial-and-error interactions with the environment over their lifetime [4]. Moreover, data inefficiency caused by trial-and-error learning mechanisms makes deep RL difficult to apply in a wide range of areas [7]. This survey aims to address these challenges by exploring recent advancements in reinforcement learning, focusing on the development of more efficient and effective learning algorithms.

The problem we address is the development of more efficient and effective reinforcement learning algorithms that can learn from trial-and-error interactions with the environment, while also being able to transfer knowledge from external expertise to facilitate the learning process [13]. Our proposed solution involves investigating recent advancements in RL, such as deep RL in computer vision [6], group-agent reinforcement learning [3], and distributed deep reinforcement learning [7]. We aim to answer the following research questions: (1) How can we improve the efficiency and effectiveness of reinforcement learning algorithms? (2) What are the key advancements in RL that can be leveraged to address the challenges faced by traditional RL paradigms?

Related work in the field of reinforcement learning includes the development of algorithms such as Q-learning, Double Q-learning, and Dueling Q-learning [5, 11, 8]. Additionally, transfer learning approaches have been explored to tackle various challenges faced by RL, by transferring knowledge from external expertise to facilitate the learning process [13]. Furthermore, recent research has focused on the development of distributed deep RL algorithms, which have shown potential in various applications such as human-computer gaming and intelligent transportation [7].

Our work differs from the existing literature in that we aim to provide a comprehensive survey of the recent advancements in reinforcement learning, focusing on the development of more efficient and effective learning algorithms.

By investigating various RL techniques and methodologies, we hope to identify key advancements that can be leveraged to address the challenges faced by traditional RL paradigms. Moreover, our survey will not only discuss the algorithms themselves but also explore their applications in various domains, providing a more in-depth understanding of the potential impact of these advancements on the AI community.

In conclusion, this survey will provide a detailed overview of recent advancements in reinforcement learning, with a focus on addressing the challenges faced by traditional RL paradigms and improving the efficiency and effectiveness of learning algorithms. By investigating various RL techniques and methodologies, we aim to identify key advancements that can be leveraged to address these challenges and contribute to the ongoing development of reinforcement learning as a powerful learning paradigm for solving sequential decision-making problems in various domains.

2 related works

Reinforcement Learning and Q-Learning Reinforcement learning is a learning paradigm for solving sequential decision-making problems, and Q-learning is one of its fundamental algorithms [13]. The Q-learning algorithm, however, is known to suffer from maximization bias, which leads to the overestimation of action values [8]. To address this issue, Double Q-learning has been proposed, which mitigates the overestimation problem but may result in slower convergence and increased memory requirements [1]. Another approach to tackle the maximization bias is Self-correcting Q-learning, which balances the overestimation and underestimation issues while maintaining similar convergence guarantees as Q-learning [8].

Deep Reinforcement Learning Deep reinforcement learning (DRL) combines reinforcement learning with deep neural networks to tackle more complex problems [6]. DRL has been successfully applied in various domains, including computer vision, where it has been used for tasks such as landmark localization, object detection, object tracking, image registration, image segmentation, and video analysis [6]. Despite its success, DRL suffers from data inefficiency due to its trial-and-error learning mechanism, leading to the development of various sample-efficient methods, such as distributed deep reinforcement learning [7].

Transfer Learning in Reinforcement Learning Transfer learning has emerged as a promising approach to address the challenges faced by reinforcement learning, such as data inefficiency, by transferring knowledge from external sources to facilitate the learning process [13]. A systematic investigation of transfer learning approaches in the context of deep reinforcement learning has been conducted, categorizing these approaches based on their goals, methodologies, compatible reinforcement learning backbones, and practical applications [13].

Policy Gradient Methods Policy gradient methods are widely used in reinforcement learning, particularly for continuous action settings. Natural policy gradients have been proposed as a more efficient alternative to traditional policy gradients, forming the foundation of contemporary reinforcement learning algorithms, such as Trust Region Policy Optimization (TRPO) and Proximal Policy Optimization (PPO) [10]. Off-policy policy gradient methods have also been developed, with the introduction of Actor Critic with Emphatic weightings (ACE), which addresses the issues of previous off-policy policy gradient methods like OffPAC and DPG [2].

Group-Agent Reinforcement Learning Group-agent reinforcement learning has been proposed as a new type of reinforcement learning problem, distinct from single-agent and multi-agent reinforcement learning [3]. In this scenario, multiple agents perform separate reinforcement learning tasks cooperatively, sharing knowledge without any cooperative or competitive behavior as a learning outcome. The Decentralised Distributed Asynchronous Learning (DDAL) framework has been introduced as the first distributed reinforcement learning framework designed for group-agent reinforcement learning, showing desirable performance and good scalability [3].

3 backgrounds

Reinforcement Learning (RL) is a learning paradigm for solving sequential decision-making problems, where an agent learns to make decisions by interacting with an environment and receiving feedback in the form of rewards or penalties [4]. The central problem in RL is to find an optimal policy, which is a mapping from states to actions, that maximizes the expected cumulative reward over time.

One of the foundational theories in RL is the concept of Markov Decision Processes (MDPs), which provide a mathematical framework for modeling decision-making problems. An MDP is defined as a tuple (S, A, P, R, γ), where S is the set of states, A is the set of actions, P is the state transition probability function, R is the reward function, and γ is the discount factor [6]. The objective in an MDP is to find a policy π that maximizes the expected return $G_t = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1}$, where t is the current time step and $\gamma \in [0, 1]$ is the discount factor that determines the importance of future rewards.

Q-learning is a popular model-free RL algorithm that estimates the action-value function $Q(s, a)$, which represents the expected return when taking action a in state s and following the optimal policy thereafter [1]. The Q-learning update rule is given by:

$$Q(s, a) \leftarrow Q(s, a) + \alpha[R(s, a) + \gamma \max_{a'} Q(s', a') - Q(s, a)], \tag{1}$$

where α is the learning rate, s' is the next state, and a' is an action in state s' [11].

Deep Reinforcement Learning (DRL) is an extension of RL that employs deep neural networks as function approximators for the value function or policy [6]. DRL has demonstrated remarkable success in various domains, including finance, medicine, healthcare, video games, robotics, and computer vision [6]. However, DRL is known to suffer from data inefficiency due to its trial-and-error learning mechanism, and several methods have been proposed to improve sample efficiency, such as environment modeling, experience transfer, and distributed modifications [3].

Policy gradient methods are another class of RL algorithms that directly optimize the policy by following the gradient of the expected return with respect to the policy parameters [9]. The policy gradient theorem provides a simplified form for the gradient, which has been widely used in on-policy learning algorithms [12]. Off-policy learning, where the behavior policy is not necessarily attempting to learn and follow the optimal policy for the given task, has been a challenging area of research, and recent work has proposed the first off-policy policy gradient theorem using emphatic weightings [2].

In summary, Reinforcement Learning aims to solve sequential decision-making problems by finding an optimal policy that maximizes the expected cumulative reward over time. Foundational theories and algorithms such as MDPs, Q-learning, DRL, and policy gradient methods provide the basis for RL research and applications in various domains [4, 6].

References

[1] D. Barber. Smoothed q-learning. *arXiv preprint arXiv:2303.08631*, 2023.

[2] M. W. Ehsan Imani, Eric Graves. An off-policy policy gradient theorem using emphatic weightings. *arXiv preprint arXiv:1811.09013*, 2018.

[3] X.-J. Z. Kaiyue Wu. Group-agent reinforcement learning. *arXiv preprint arXiv:2202.05135*, 2022.

[4] C. Li. Some insights into lifelong reinforcement learning systems. *arXiv preprint arXiv:2001.09608*, 2020.

[5] M. G. P. E. Li Meng, Anis Yazidi. Expert q-learning: Deep reinforcement learning with coarse state values from offline expert examples. *arXiv preprint arXiv:2106.14642*, 2021.

[6] K. Y. K. L. M. S. Ngan Le, Vidhiwar Singh Rathour. Deep reinforcement learning in computer vision: A comprehensive survey. *arXiv preprint arXiv:2108.11510*, 2021.

[7] S. S. J. Y. M. Z. K. H. B. L. L. W. Qiyue Yin, Tongtong Yu. Distributed deep reinforcement learning: A survey and a multi-player multi-agent learning toolbox. *arXiv preprint arXiv:2212.00253*, 2022.

[8] M. R. Rong Zhu. Self-correcting q-learning. *arXiv preprint arXiv:2012.01100*, 2020.

[9] J. Schmidt. Morphisms of networks of hybrid open systems. *arXiv preprint arXiv:1911.09048*, 2019.

[10] W. J. A. van Heeswijk. Natural policy gradients in reinforcement learning explained. *arXiv preprint arXiv:2209.01820*, 2022.

[11] J. L. Wei Liao, Xiaohui Wei. Smooth q-learning: Accelerate convergence of q-learning using similarity. *arXiv preprint arXiv:2106.01134*, 2021.

[12] V. K. Yemi Okesanjo. Revisiting stochastic off-policy action-value gradients. *arXiv preprint arXiv:1703.02102*, 2017.

[13] A. K. J. J. Z. Zhuangdi Zhu, Kaixiang Lin. Transfer learning in deep reinforcement learning: A survey. *arXiv preprint arXiv:2009.07888*, 2020.