A Survey of Deep Reinforcement Learning Based on Multi-Particle Environments

Chuhao Weng *

Department of Computing, King's College London, London, United Kingdom

* Corresponding Author Email: k22056097@kcl.ac.uk

Abstract. This paper extensively explores the application of deep reinforcement learning within multi-agent environments, a domain marked by intricate challenges and burgeoning opportunities. Multi-agent environments inherently entail the presence of multiple interacting intelligent agents, thereby amplifying the complexity of decision-making and control. Traditional reinforcement learning methods, when confronted with these intricate settings, manifest limitations that have prompted a compelling shift towards the fusion of deep learning and reinforcement learning. The journey commences by elucidating the multifaceted challenges and intricacies pervasive in multi-agent environments. It systematically delineates the constraints of conventional reinforcement learning techniques, underscoring their inadequacies in effectively addressing the multifarious complexities presented by these environments. As the narrative unfolds, it places a spotlight on the transformative potential of deep learning in surmounting the formidable challenges associated with decision-making in multi-agent settings. A critical examination ensues, encompassing a thorough review of pioneering applications of methodologies such as value functions, policy gradients, and actor-critic approaches within the realm of multi-agent systems. Through an exhaustive exploration of existing literature, this article endeavors to unveil the vanguard developments and persisting challenges within the domain of deep reinforcement learning in multi-agent environments. Furthermore, this paper conducts a comparative analysis of the performance of two prominent algorithms, namely Q-learning and the Deep Q Network (DQN) algorithm. This analysis serves as a compass, elucidating the trajectory of development within the sphere of deep reinforcement learning. Ultimately, this paper acts as a compass, guiding future research and innovation in the dynamic and promising field of deep reinforcement learning within multi-agent environments.

Keywords: Multi-agent Particle Environment; Q-learning; DQN; Deep Reinforcement Learning.

1. Introduction

Deep reinforcement learning (DRL) stands as a pivotal research domain within artificial intelligence, capturing extensive attention and research efforts. It offers a powerful paradigm for enabling intelligent agents to make optimal decisions in complex, dynamically evolving environments. Multi-agent scenarios, in particular, introduce numerous intricate challenges, where agents must navigate both the dynamic environment and the interactions and dependencies among fellow agents. In multi-agent settings, conventional reinforcement learning methods often prove inadequate, falling short of delivering satisfactory outcomes [1]. The complexities inherent in these environments necessitate more sophisticated decision-making capabilities, motivating researchers to explore the fusion of deep learning and reinforcement learning. This integration holds the promise of unlocking new horizons in decision-making and control within multi-agent systems.

This article primarily aims to provide a comprehensive examination of the potential and challenges within deep reinforcement learning research in multi-agent contexts, devoid of any first or second-person perspectives [2]. To achieve this goal, it commences by elucidating the unique challenges and intricacies posed by multi-agent environments. It delineates the limitations of traditional reinforcement learning approaches when confronted with such intricacies, underscoring the imperative need for advanced methodologies. Subsequently, it delves into the pivotal role of deep learning in advancing the field of reinforcement learning, focusing on the progress made in addressing the formidable challenges associated with decision-making in multi-agent settings. This encompasses a nuanced exploration of classical and state-of-the-art research endeavors, with a specific emphasis

on innovative applications of key DRL techniques like value functions, policy gradients, and actor-critic approaches within the domain of multi-agent systems. Through a comprehensive survey of existing literature, the objective is to unveil the forefront of advancements and hurdles in the realm of DRL within multi-agent environments [3]. The ambition is to provide clarity regarding the trajectory of DRL development, elucidating the distinctions between foundational algorithms such as Q-learning and modern approaches like the DQN algorithm. In doing so, this article aims to contribute to a deeper understanding of the potential of DRL in multi-agent scenarios and to inspire future research endeavors in this dynamic and promising field.

2. Related Work

In recent years, substantial progress has been witnessed in the domain of DRL within multi-agent environments. Researchers have diligently explored innovative methodologies to grapple with the intricacies that emerge from interactions among multiple agents in dynamic settings. One noteworthy research avenue revolves around the extension of conventional reinforcement learning algorithms, such as Q-learning and policy gradient methods, to the multifaceted realm of multi-agent scenarios. A case in point is the application of algorithms rooted in deep reinforcement learning to dynamic pre-deployment strategies for Unmanned Aerial Vehicles [4]. An alternative direction of investigation entails harnessing advanced deep learning techniques to elevate the performance of multi-agent systems. An exemplary contribution to this domain is the Multi-Agent Soft Actor-Critic algorithm, introduced by Xiao Shuo et al. This innovative approach introduces centralized training and communication mechanisms aimed at fostering cooperation and informed decision-making among agents while mitigating the risks of competition and conflicts. The outcome is a marked improvement in performance when compared to traditional multi-agent systems [5].

Moreover, researchers have embarked on an extensive exploration of multi-agent reinforcement learning's applications across various domains, spanning collaborative robotics to economic contexts. Pioneering strategies like hierarchical control collaborative navigation have been propounded, ushering multi-agent deep reinforcement learning into collaborative navigation scenarios [6]. The realm of industrial wireless network edge-side collaborative resource allocation has witnessed the introduction of the Multi-Agent Deep Reinforcement Learning-Based Resource Allocation algorithm, a novel resource allocation mechanism underpinned by multi-agent deep reinforcement learning [7]. Additionally, spectrum-sharing algorithms grounded in multi-agent deep reinforcement learning have been devised for the establishment of vehicular networks [8]. Nevertheless, the pursuit of stable and efficient learning in complex multi-agent environments persists as a formidable challenge. The dynamic nature of interactions among agents continues to introduce complications linked to non-stationarity. Tackling the intricate balance between centralized and decentralized learning, managing scenarios with partial observability, and adapting to contexts featuring a high number of agents remain active areas of exploration [9]. To encapsulate, extant research serves as a testament to substantial headway in the application of deep reinforcement learning techniques within multi-agent environments. Nevertheless, the persisting challenges and ongoing research endeavors underscore the imperative necessity for innovative methodologies to effectively grapple with the intricacies arising from interactions among multiple agents [10].

3. Methodology

3.1. Q-Learning Algorithm

The Q-learning algorithm is a foundational method in reinforcement learning, primarily employed to tackle Markov Decision Process (MDP) problems. At its core, Q-learning endeavors to construct a Q-value function, an estimator of cumulative reward values (Q-values) for each conceivable state-action pair. In essence, Q-values symbolize the expected cumulative rewards obtainable by taking specific actions within given states. This algorithm dynamically updates the Q-value function, with

the aim of learning the optimal policy. Consequently, the agent selects actions associated with the highest Q-values in each state to maximize cumulative rewards. Mathematically, the Q-learning update rule can be expressed as Formula 1:

$$Q(s, a) = (1 - \alpha) * Q(s, a) + \alpha * [R + \gamma * \max(Q(s', a'))] \quad (1)$$

Q-learning's versatility extends to multi-agent environments, where each agent operates as an independent entity, equipped with its unique state and action space. In multi-agent scenarios, while agents may share information, Q-learning enables them to autonomously learn their respective Q-value functions. This characteristic greatly facilitates decision-making in the intricate landscapes of multi-agent systems. Despite its utility, Q-learning faces distinct challenges in multi-agent settings:

1. Cooperation and Competition: Multi-agent environments often encompass both cooperative and competitive aspects, necessitating agents to collaborate or compete to accomplish tasks. Q-learning can effectively manage these dynamics by independently learning Q-value functions for each agent.
2. Exploration Challenge: The expansive state space of multi-agent environments can make a thorough exploration of all possible state-action pairs impractical. Strategies like the ε-greedy approach or other exploration techniques are deployed to address this challenge.
3. Information Sharing: In specific scenarios, agents in multi-agent environments may need to share information to achieve common objectives. One approach involves utilizing centralized training and distributed execution, allowing agents to exchange information as needed.

Q-learning stands as a robust reinforcement learning algorithm applicable to multi-agent environments. The schematic diagram of the Q-learning algorithm is depicted in Fig. 1. Its approach of treating each agent as an independent entity, enabling them to independently learn their Q-value functions, facilitates effective decision-making and task-solving in these intricate settings. However, it is essential to consider exploration and information-sharing challenges when implementing Q-learning in multi-agent environments.

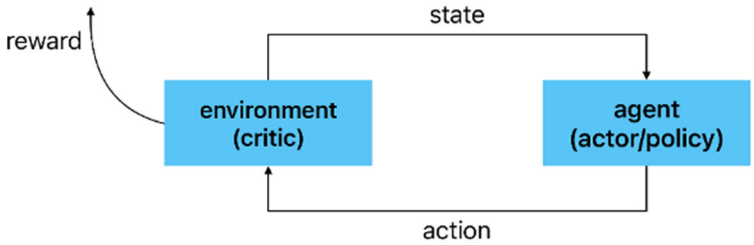

Fig. 1 Q-learning schematic diagram (Photo/Picture credit: Original)

3.2. DQN Algorithm

The DQN algorithm represents a fusion of deep learning and reinforcement learning, tailor-made for addressing MDP challenges. The schematic diagram illustrating the DQN algorithm is presented in Fig. 2. At its core, DQN employs deep neural networks to estimate the Q-value function, effectively extending the Q-value table of traditional Q-learning to accommodate continuous or high-dimensional state spaces. Here's an overview of the DQN process:

1. Neural Network Training: Train a deep neural network with states as inputs and Q-value estimates for each action as outputs.
2. Action Selection: Choose actions, typically employing an ε-greedy strategy, where ε regulates the balance between exploration and exploitation.
3. Execution and Observation: Execute the selected action, observe the reward, and record the next state.

4. Experience Replay: Store experience tuples (state, action, reward, next state) in an experience replay buffer.
5. Batch Sampling: Randomly sample batches of experiences from the replay buffer for training the neural network.
6. Loss Calculation and Weight Update: Calculate the loss using fixed Q-targets and update the neural network's weights via gradient descent.
7. Repeat: Iterate through the above steps until convergence or a predefined number of training steps are reached.

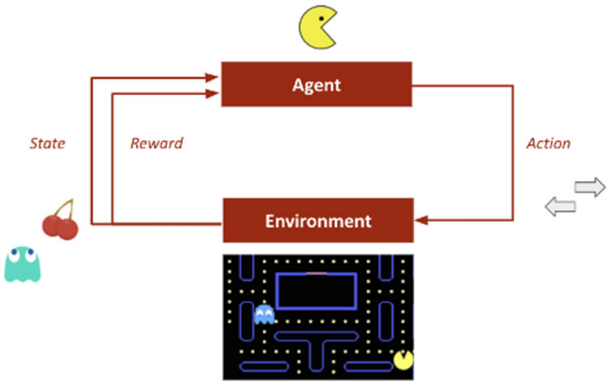

Fig. 2 DQN schematic diagram (Photo/Picture credit: Original)

DQN's versatility extends to multi-agent environments, where each agent operates as an independent entity with its unique DQN for estimating the Q-value function. Agents make action choices based on their states and individual Q-value functions. In multi-agent scenarios, DQN adeptly handles cooperation and competition by allowing agents to independently learn Q-value functions. Nevertheless, DQN encounters specific challenges in multi-agent environments.

1. Exploration Challenge: Multi-agent environments often entail expansive state spaces, making comprehensive exploration arduous. The ε-greedy strategy, proven effective in DQN, can also be applied to multi-agent contexts.
2. Information Sharing: In certain situations, multi-agents must share information to achieve shared objectives. One approach is to employ centralized training and distributed execution, enabling agents to exchange pertinent information.

3.3. Comparison

DQN offers several advantages over traditional Q-learning in multiple aspects:

1. Handling High-Dimensional State Spaces: Q-learning relies on a Q-value table, which becomes impractical for large or continuous state spaces. DQN, with its deep neural networks, adeptly manages high-dimensional state spaces, making it suitable for complex data like images and sounds.
2. Generalization Ability: DQN learns general patterns among states, facilitating better generalization capabilities. It can make informed decisions even in states it hasn't previously encountered.
3. Continuous Action Spaces: Unlike Q-learning, which handles discrete actions exclusively, DQN accommodates continuous action spaces. It can produce continuous action values, rendering it suitable for diverse problems requiring continuous control.
4. Experience Replay: DQN's experience replay stores and randomly samples past experiences, stabilizing training and diminishing sample correlation. This mitigates the risk of becoming stuck in local optima and enhances learning efficiency.
5. Target Network: DQN introduces target networks, enhancing training stability and reducing fluctuations in Q-value targets during training. This contributes to more consistent learning and quicker convergence.

6. Deep Learning Advantages: DQN leverages the inherent strengths of deep learning to autonomously extract features and representations. This eliminates the need for manual feature engineering, rendering DQN exceptionally effective in handling raw perceptual data like images.

7. Applicability: DQN boasts wide applicability across various reinforcement learning domains, spanning Atari games, robot control, autonomous driving, and more. It has achieved remarkable success in multiple domains.

However, despite DQN's myriad advantages, it does present certain challenges and limitations. Longer training times and sensitivity to hyperparameters are among the challenges. Additionally, for specific problems, traditional Q-learning methods may still demonstrate competitiveness. Therefore, the selection between DQN and Q-learning should be guided by the particular problem and its requirements.

4. Results and Analysis

In the realm of multifunctional radar (MFR) operations, the demand for efficient cognitive interference decision-making has surged. However, traditional Q-learning approaches have begun to exhibit diminishing efficiency as the number of tasks performed by MFRs continues to rise, as observed in previous research1. To address this challenge, researchers have turned their attention toward employing DQN for MFR cognitive interference decision-making. In this section, we present a comparative analysis of experimental results to evaluate the performance of DQN against traditional Q-learning. Experimental findings reveal a noteworthy trend: as the number of radar tasks increases, the traditional Q-learning algorithm experiences a substantial escalation in the time required for decision-making. This phenomenon becomes particularly pronounced when the number of radar tasks reaches significant levels. Fig. 3 illustrates the performance comparison between Q-learning and DQN. The x-axis denotes the total number of radar tasks in units, while the y-axis represents the decision completion time in seconds. The blue line represents the performance of the Q-learning algorithm, and the red line represents the performance of the DQN algorithm.

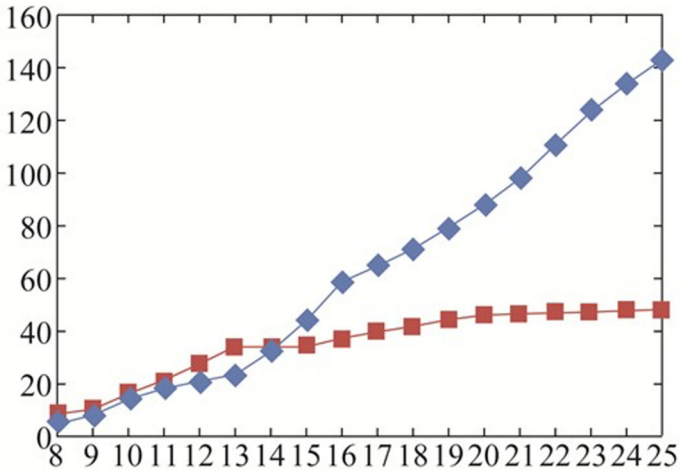

Fig. 3 Q-learning and DQN performance comparison chart (Photo/Picture credit: Original)

In sharp contrast, the utilization of DQN for MFR yields considerably higher decision-making efficiency across varying task numbers. Notably, when dealing with a large volume of tasks, the DQN algorithm outperforms the Q-learning algorithm quite conspicuously. This superiority chiefly manifests in the efficiency of decision-making. Moreover, this advantage becomes more pronounced as the task count increases. The research results distinctly illustrate that when it comes to handling cognitive interference decisions within the ambit of multifunction radar tasks, the DQN-based approach significantly outperforms the traditional Q-learning method in terms of decision-making efficiency. This disparity is particularly pronounced when dealing with a substantial number of tasks.

These findings underscore the superior performance and efficiency of DQN in addressing cognitive interference issues within complex multi-task environments.

The transition from traditional Q-learning to the DQN-based approach marks a pivotal advancement in cognitive interference decision-making for multifunctional radar systems. This transition not only enhances decision-making efficiency but also demonstrates adaptability in addressing the escalating demands of modern multifunction radar tasks. As multifunctional radar systems continue to evolve, the utilization of DQN promises to play a crucial role in optimizing their performance and ensuring effective cognitive interference management. This research paves the way for the integration of DQN-based approaches into multifunctional radar systems, offering improved decision-making capabilities and bolstering their effectiveness across various operational scenarios.

5. Conclusion

The domain of deep reinforcement learning in multi-agent environments presents a dynamic landscape brimming with challenges and opportunities. This paper has meticulously examined the multifaceted challenges inherent in multi-agent settings, encompassing cooperation, competition, exploration dilemmas, and information-sharing intricacies. In the methodology section, we provided a comprehensive insight into the mechanisms of both the Q-learning and DQN algorithms, elucidating their applicability within multi-agent environments. Through a rigorous comparative analysis of experimental outcomes, a significant revelation emerged: as the number of tasks escalates, the decision-making time associated with the Q-learning algorithm experiences a noteworthy upsurge. In stark contrast, the DQN algorithm, customized for MFR tasks, demonstrates remarkable decision-making efficiency, irrespective of the task quantity. This compelling finding underscores the superior performance of DQN when it comes to tackling cognitive interference concerns within the intricate realm of multi-task environments.

The integration of deep reinforcement learning into multi-agent environments heralds a promising approach for resolving decision-making challenges within complex multi-agent systems. The evident potential of this approach is further substantiated by its practical applications. However, it is essential to acknowledge that this field is characterized by its own set of challenges. It is imperative that future research endeavors continue to delve into the exploration of novel deep reinforcement learning methods to effectively address the myriad challenges posed by multi-agent environments. In conclusion, while the application of deep reinforcement learning in multi-agent environments holds vast potential, there is still a pressing need for ongoing research and innovation to surmount existing challenges. It is our hope that this review serves as a valuable resource, offering insights and guidance to inspire and inform future research in this exciting and evolving field.

References

[1] Tang Lun, et al. "Dynamic pre-deployment strategy for UAVs based on multi-agent deep reinforcement learning." Journal of Electronics and Information Technology 45.6 (2023): 2007-2015.

[2] Liu Jian, et al. "Breast cancer pathogenic gene prediction based on multi-agent reinforcement learning." Acta Automata 48.5 (2022): 1246-1258.

[3] Cong Z, Xin-yuan Z, Xing-hua L I, et al. Intelligent Delay Matching Method for Parking Allocation System via Multi-Agent Deep Reinforcement Learning [J]. China Journal of Highway and Transport, 2022, 35(7): 261.

[4] Wei S, Yang-He F, Guang-Quan C, et al. Research on multi-aircraft cooperative air combat method based on deep reinforcement learning [J]. Acta Automatica Sinica, 2021, 47(7): 1610-1623.

[5] LIU, Xiaoyu, et al. "End-edge collaborative resource allocation in industrial wireless networks based on multi-agent deep reinforcement learning." Frontiers 23.1 (2022): 47-60.

[6] Zhu F, Wu W, Fu Y, et al. A dual deep network based secure deep reinforcement learning method [J]. Chinese Journal of Computers, 2019, 42(8): 1812-1826.

[7] Xiao Shuo, et al. "multi-agent deep reinforcement learning algorithm based on SAC." Journal of Electronics 49.9 (2021): 1675.

[8] Wang Weinian, et al. "Internet of Vehicles spectrum sharing based on multi-agent deep reinforcement learning." Journal of Electronics (2023): 1.

[9] Sun Changyin, and Mu Chaoxu. "Several key scientific issues in multi-agent deep reinforcement learning." Journal of Automation 46.7 (2020): 1301-1312.

[10] Zhang Baikai, and Zhu Weigang. "DQN cognitive interference decision-making method for multi-function radar." Systems Engineering and Electronic Technology 42.4 (2020): 819-825.