Deep Reinforcement Learning A brief survey

Ambati Praggnya sai
640: Introduction to AI
Semester II
M03543484
pa752s@missouristate.edu

Neha Singuluri
640: Introduction to AI
Semester II
M03543630
ns9392s@missouristate.edu

I. INTRODUCTION

Deep reinforcement learning (DRL) represents a cutting-edge approach in artificial intelligence, empowering autonomous systems with a profound understanding of the visual world . Through interaction and trial-and-error, DRL allows agents to learn optimal behaviors by fusing deep neural networks with reinforcement learning. This survey demonstrates DRL's capacity to solve issues that were previously unsolvable by examining its applications, underlying theories, and effects. DRL's adaptability is clear—it may be used to operate real-world robotics or to master video games. In order to provide readers with an understanding of how artificial intelligence is developing and what the future holds for deep reinforcement learning, this paper explores fundamental DRL ideas, important algorithms, hierarchical learning, intrinsic motivation, imitation learning, and current research trends.

II. BACKGROUND ON DEEP REINFORCEMENT LEARNING

Through interactions with their surroundings, Deep Reinforcement Learning (DRL), a novel approach that combines reinforcement learning and deep learning, enables machines to learn complex behaviors and decision-making processes [1]. Without the need for human feature engineering, deep learning (DRL) allows agents to autonomously learn hierarchical representations of high-dimensional data, like pictures or sensor inputs [1,2] . The combination of deep learning methods and reinforcement learning algorithms has resulted in remarkable advancements across various domains, such as surpassing human performance in video games, improving robotics capabilities, promoting natural language processing, and facilitating advancements in autonomous driving systems. DRL agents learn optimal strategies by receiving feedback in the form of rewards or penalties based on their actions, allowing them to refine their decision-making abilities over time through iterative learning. Although its remarkable accomplishments, DRL still confronts a number of difficulties that motivate ongoing study, including sample efficiency, applicability to new settings, and the interpretability of learnt policies [3]. DRL has an impact on a wide range of sectors, including autonomous systems, healthcare, and finance, where intelligent and adaptable agents may pick up on environmental changes and change with them. All things considered, deep reinforcement learning (DRL) represents a revolutionary technique that has the ability to completely change artificial intelligence and usher in a new era of intelligent, autonomous systems that are able to solve challenging problems and make complex decisions.

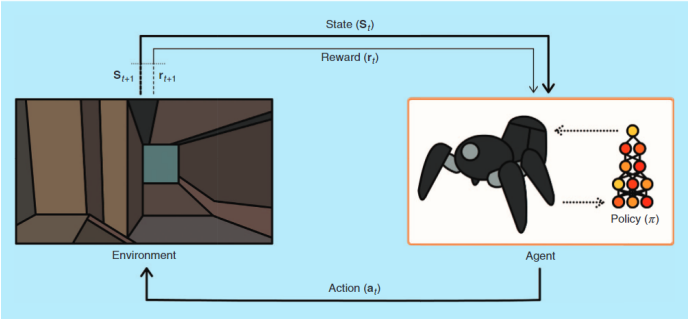

Fig. 1. The perception-action-learning loop.

III. TECHNIQUES IN DEEP REINFORCEMENT LEARNING

1) Value-Based Methods: Approximating the ideal action-value function, which links state-action pairings to expected cumulative rewards, is the goal of value-based techniques like Deep Q-Networks (DQN) [3] . DQN allows agents to learn complicated decision-making tasks directly from raw sensory inputs by estimating Q-values using deep neural networks [3] . More effective and reliable training is made possible by methods like experience replay stores and randomly sampled transitions from the agent's prior experiences. By offering defined targets for Q-value estimation, target networks—copies of the main network that are updated with the main network's parameters on a regular basis—stabilize training and avoid damaging feedback loops during learning.

2) Policy-Based Methods: The policy function, which describes the probability distribution over actions given a state, is directly optimized by policy-based techniques such as Trust Region Policy Optimization (TRPO) and Proximal Policy Optimization (PPO)[4]. To update the policy parameters in the direction of higher predicted rewards, these strategies make use of policy gradient techniques. Actor-critic architectures provide more effective learning of optimal policies by combining policy improvement (actor) with value estimate (critic). These approaches handle continuous action spaces and offer reliable and sample-efficient learning by directly optimizing the policy.

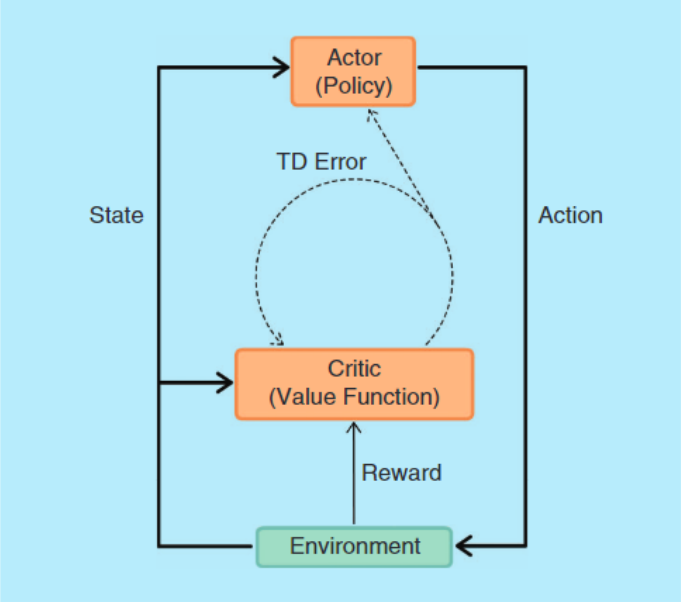

Fig. 2. The actor-critic setup.

3) Model-Based Reinforcement Learning: In order to increase sampling efficiency and planning skills, model-based reinforcement learning entails learning a model of the environment,[5] such as transition dynamics or reward function. Without physically interacting with the environment, agents can simulate potential future trajectories and make well-informed judgments by using trained models. Decision-making is improved through integration with deep neural networks,[6] which facilitates the learning of intricate and nonlinear environment models. The exploration-exploitation trade-off is addressed by model-based approaches, which more effectively direct the agent towards high-reward areas by offering more precise forecasts of future states and rewards.

4) Exploration Strategies: Different exploration tactics are used in DRL to handle trade-offs between exploration and exploitation. To promote exploration and progressively take use of learnt rules, epsilon-greedy exploration chooses random actions with a given probability, or epsilon [7]. SoftMax exploration prioritizes activities with larger expected rewards by assigning probabilities to them based on their estimated values. By adding unpredictability to the process of choosing an action,[8] noise injection promotes experimentation in uncharted territory. [9] By striking a balance between the exploration of novel states and actions and the application of previously acquired knowledge, these techniques influence the learning process and ultimately result in the identification of the best courses of action in unpredictable contexts.

IV. Applications of Deep Reinforcement Learning

1) Recommendation Systems: Recommendation systems use DRL to enhance user engagement and personalize content. [10] Algorithms can improve user experience by suggesting suitable items, movies, or articles based on user behavior and preference data.

2) Natural Language Processing (NLP): Text generation, sentiment analysis, and language translation are NLP activities where deep learning is used. More natural human-computer interactions are made possible by agents' ability to respond to discussions in a logical and contextually appropriate manner. [11]

3) Supply Chain Management: By forecasting changes in demand, managing inventory levels, and dynamically allocating resources, DRL enhances supply chain operations. It aids companies in cutting expenses, increasing productivity, and reacting fast to shifts in the market.[11]

4) Virtual Assistants and Chatbots: DRL enables chatbots and virtual assistants to comprehend user inquiries, provide contextually appropriate answers, and help users with a range of tasks. It makes it possible for people and robots to interact with more intelligence and creativity.

5) Simulation and Gaming Industry: DRL is used to enhance virtual environments, create more realistic non-player characters (NPCs), and improve the overall gaming experience.

V. Challenges

A major problem in deep reinforcement learning (DRL) is sample efficiency because many algorithms require a large number of samples to learn well, which can be problematic in contexts with limited resources[13]. Getting the models trained in simulations to perform well in real-world circumstances is similarly difficult because of differences in the environments. Ensuring safety, particularly in vital domains such as autonomous vehicles and healthcare, necessitates robust algorithms that can manage ambiguities and hostile inputs [14,15]. It is essential to strike a balance between exploration and exploitation, which encourages the development of new tactics. Reaching these obstacles is essential to extending DRL's usefulness and reach into other sectors.

VI. Future Work

In considering future directions for Deep Reinforcement Learning (DRL) , it is essential to highlight potential avenues for advancement and research in this dynamic field [16]. Subsequent investigations in DRL may concentrate on creating new methods and algorithms that improve sample efficiency, enabling agents to discover ideal policies with less interactions with the surroundings [17] . Investigating techniques like curriculum learning, transfer learning, and meta-learning may result in considerable gains in learning performance and speed across a range of tasks and domains. In order to tackle the problem of generalization in DRL, future research endeavors can delve into methods that facilitate agents' efficient generalization to unfamiliar tasks and contexts.The focus of research could be on creating algorithms that enhance the resilience, flexibility, and transferability of learnt policies so that DRL systems can function dependably under a variety of settings in real-world scenarios [18] . Giving top priority to the creation of understandable and interpretable models that

shed light on autonomous agents' decision-making procedures could improve accountability, transparency, and confidence in AI applications—especially in areas where safety is at stake [19] . Examining multi-task and meta-learning techniques in DRL may create new opportunities for agents to effectively learn and develop their abilities in a variety of related tasks. In complex and dynamic situations, DRL systems can achieve better generalization, faster learning, and adaptive behavior by utilizing shared knowledge and experiences from many tasks.In order to ensure the responsible development and application of AI technologies for the benefit of society, future directions in DRL should also take into account the ethical and social implications of autonomous learning systems. Specifically, [20] they should address concerns about bias, fairness, privacy, and accountability in DRL algorithms.

VII. Conclusion

In conclusion, the field of deep reinforcement learning is experiencing a transformative phase, driven by advancements in policy search methods, evolutionary algorithms, advantage function learning, and strategies to enhance sample efficiency. The current research efforts aim to combine deep neural networks with reinforcement learning algorithms in order to address complex tasks and enhance learning effectiveness. Although there has been significant advancement, ongoing difficulties such as the complexity of samples, the capacity to apply knowledge to new situations, and the ability to understand and explain results highlight the necessity for more investigation. The combination of deep learning and reinforcement learning holds the potential to create autonomous agents that possess a comprehensive comprehension of the visual environment. Additionally, this integration represents a substantial advancement towards the achievement of sophisticated artificial intelligence systems. Researchers want to develop autonomous systems that can navigate and interact with complicated surroundings in a way that is similar to human cognition by combining the capability of various technologies. The continued exploration of these avenues for study not only advances the discipline but also paves the way for revolutionary developments in deep reinforcement learning, which will influence autonomous systems and artificial intelligence in the years to come.

References

[1] K. Kansky, T. Silver, D. A. Mély, M. Eldawy, M. Lázaro-Gredilla, X. Lou, N. Dorfman, S. Sidor, S. Phoenix, and D. George, "Schema networks: zero-shot transfer with a generative causal model of intuitive physics," in Proc. Int. Conf. Machine Learning, 2017, pp. 1809–1818.

[2] K. Kansky, T. Silver, D. A. Mély, M. Eldawy, M. Lázaro-Gredilla, X. Lou, N. Dorfman, S. Sidor, S. Phoenix, and D. George, "Schema networks: zero-shot transfer with a generative causal model of intuitive physics," in Proc. Int. Conf. Machine Learning, 2017, pp. 1809–1818.

[3] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: a review and new perspectives," IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 8, pp. 1798–1828, 2013.

[4] P. Christiano, Z. Shah, I. Mordatch, J. Schneider, T. Blackwell, J. Tobin, P. Abbeel, and W. Zaremba. (2016). Transfer from simulation to real world through learn- ing deep inverse dynamics model. arXiv. [Online]. Available: https://arxiv.org/ abs/1610.03518

[5] J. Foerster, Y. M. Assael, N. de Freitas, and S. Whiteson, "Learning to communi- cate with deep multi-agent reinforcement learning," in Proc. Neural Information Processing Systems, 2016, pp. 2137–2145

[6] F. Gomez and J. Schmidhuber. "Evolving modular fast-weight networks for con- trol," in Proc. Int. Conf. Artificial Neural Networks, 2005, pp. 383–389.

[7] V. Mnih, A. P. Badia, M. Mirza, A. Graves, T. P. Lillicrap, T. Harley, D. Silver, and K. Kavukcuoglu, "Asynchronous methods for deep reinforcement learn- ing," in Proc. Int. Conf. Learning Representations, 2016

[8] B. C. Stadie, S. Levine, and P. Abbeel, "Incentivizing exploration in reinforcement learning with deep predictive models," in NIPS Workshop on Deep Reinforcement Learning, 2015

[9] O. Nachum, M. Norouzi, K. Xu, and D. Schuurmans. (2017). Bridging the gap between value and policy based reinforcement learning. arXiv. [Online]. Available: https://arxiv.org/abs/1702.08892

[10] A. Y. Ng, A. Coates, M. Diel, V. Ganapathi, J. Schulte, B. Tse, E. Berger, and E. Liang, "Autonomous inverted helicopter flight via reinforcement learning," in Proc. Int. Symp. Experimental Robotics, 2006, pp. 363–372.

[11] J. Oh, X. Guo, H. Lee, R. L. Lewis, and S. Singh, "Action-conditional video prediction using deep networks in Atari games," in Proc. Neural Information Processing Systems, 2015, pp. 2863–2871.

[12] C. Tessler, S. Givony, T. Zahavy, D. J. Mankowitz, and S. Mannor, "A deep hier- archical approach to lifelong learning in Minecraft," in Proc. Association for the Advancement Artificial Intelligence, 2017, pp. 1553–1561

[13] E. Tzeng, C. Devin, J. Hoffman, C. Finn, X. Peng, S. Levine, K. Saenko, and T. Darrell, "Towards adapting deep visuomotor representations from simulated to real environments," in Workshop Algorithmic Foundations Robotics, 2016.

[14] H. van Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with double Q-learning," in Proc. Association for the Advancement of Artificial Intelligence, 2016, pp. 2094–2100.

[15] T. Salimans, J. Ho, X. Chen, and I. Sutskever. (2017). Evolution strategies as a scalable alternative to reinforcement learning. arXiv. [Online]. Available: https:// arxiv.org/abs/1703.03864.

[16] T. Schaul, D. Horgan, K. Gregor, and D. Silver, "Universal value function approximators," in Proc. Int. Conf. Machine Learning, 2015, pp. 1312–1320.

[17] M. Wulfmeier, P. Ondruska, and I. Posner, "Maximum entropy deep inverse reinforcement learning," in NIPS Workshop on Deep Reinforcement Learning, 2015.

[18] D. J. Rezende, S. Mohamed, and D. Wierstra, "Stochastic backpropagation and approximate inference in deep generative models," in Proc. Int. Conf. Machine Learning, 2014, pp. 1278–1286

[19] J. Oh, X. Guo, H. Lee, R. L. Lewis, and S. Singh, "Action-conditional video prediction using deep networks in Atari games," in Proc. Neural Information Processing Systems, 2015, pp. 2863–2871.

[20] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," in Proc. Int. Conf. Learning Representations, 2016.