Applications of Deep Reinforcement Learning in Other Industries

Subjects: Computer Science, Artificial Intelligence
Contributor: Xuanchen Xiang

Reinforcement Learning (RL) is an approach to simulate the human's natural learning process, whose key is to let the agent learn by interacting with the stochastic environment.

Keywords: reinforcement learning ; deep reinforcement learning

1. Introduction

DRL is the combination of Deep Learning and Reinforcement Learning, and it's more robust than Deep Learning or Reinforcement Learning. However, it inherits some drawbacks that DP and RL have.

Deep Learning extracts features and tasks from data. Generally, the more data provided in training, the better performance DL has. Deep Learning requires lots of data and high-performance GPUs to achieve specific functions. Due to the complex data models, it's costly to train the models. There's no standard rule for selecting DL tools or architectures, and tuning the hyperparameters could also be time-consuming. This makes DL unpractical in many domains.

Reinforcement Learning imitates the learning process of humans. It is trained by making and then avoiding mistakes. It can solve some problems that conventional methods can't solve. In some tasks, it also has the ability to surpass humans. However, RL also has some limitations. First of all, too much reinforcement might cause an overload of states, diminishing the results. Secondly, RL assumes the environment is a Markovian model, in which the probability of the event depends only on the previous state. Thirdly, it has the disadvantages of the curse of dimensionality and the curse of real-world samples. What's more, we have mentioned the challenges of setting up rewards, balancing exploration and exploitation, etc. [1]. Reinforcement Learning is an expensive and complex method, so it's not preferable for simple tasks.

Employing DRL in the real world is complex. Dulac-Arnold et al. [2] addressed nine significant challenges of practical RL in the real world. They presented examples for each challenge and provided some references for deploying RL:

- Modeling the real world is complex. Many systems cannot be directly trained on. An off-line off-policy approach [2] could be deployed to replace a previous control system. Logs from the policy are available, and the policy is trained with batches of data obtained from the control algorithm.
- Practical systems do not have separate training and evaluation environments. The agent must explore and act reasonably and safely. Thus, a sample-efficient and performant algorithm is crucial. Finn et al. [3] proposed Model Agnostic Meta-Learning (MAML) to learn within a distribution with few shot learning. Osband et al. [4] used Bootstrapped DQN to learn an ensemble of Q-networks and Thompson Sampling to achieve deep efficient exploration. Using expert demonstrations to bootstrap the agent can also improve efficiency, which has been combined with DQN [5] and DDPG [6].
- Real-world environments usually have massive and continuous state and action spaces. Dulac-Arnold et al. [7] addressed the challenge for sizeable discrete action spaces. Action-Elimination Deep Q-Network (AE-DQN) [8] and Deep Reinforcement Relevance Network (DRRN) [9] also deals with the issue.
- The learned policy might violate the safety constraints. Constrained MDP (CMDP) [2] and budgeted MDP [10] take the constraint components into consideration during training.
- Considering POMDP problems, Dulac-Arnold et al. [2] presented Robust MDPs, where the learned policy maximizes the worst-case value function.

- Formulating multi-dimensional reward functions is usually necessary and complicated. Distributional DQN Bellemare et al. [11] models the percentile distribution of the rewards. Dulac-Arnold et al. [2] presented multi-objective analysis and formulated the global reward function as a linear combination of sub-rewards. Abbeel and Ng [12] gave an algorithm is based on inverse RL to try to recover the unknown reward function.

- Policy explainability is vital for real-world policies as humans operate the systems.

- Policy inference should be made in real-time at the control frequency of the system. Hester et al. [13] presented a parallel real-time architecture for model-based RL. AlphaGo [14] improves with more rollouts rather than running at a specific frequency.

- Most natural systems have delays in the perception of the states, the actuators, or the return. Hung et al. [15] proposed a memory-based algorithm where agents use recall of memories to credit actions from the past. Arjona-Medina et al. [16] introduced RUDDER (Return Decomposition for Delayed Rewards) to learn long-term credit assignments for delayed rewards.

2. Applications

2.1. Transportation

An intelligent transportation system (ITS) [17] is an application that aims to provide safe, efficient, and innovative services to transport and traffic management and construct more intelligent transport networks. The technologies include car navigation, traffic signal control systems, container management systems, variable message signs, and more. Effective technologies like sensors, Bluetooth, radar, etc., have been applied in ITS and have been widely discussed. In recent years, with DRL steps into vision, the application of DRL in ITS has been researched. Haydari and Yilmaz [18] presented a comprehensive survey on DRL for ITS.

2.2. Industrial Applications

2.2.1. Industry 4.0

Industry 4.0, which denotes The Fourth Industrial Revolution, uses modern innovative technology to automate traditional manufacturing and industrial practices. Artificial intelligence enables many applications in Industry 4.0, including predictive maintenance, diagnostics, and management of manufacturing activities and processes [19].

Robotics, including manipulation, locomotion, etc., will prevail in all aspects of industrial applications, which was mentioned in [1]. For example, Schoettler et al. [20] discussed insertion tasks, particularly in industrial applications; Li et al. [21] also discussed a skill-acquisition DRL method to make robots acquire assembly skills.

Inspection and Maintenance

Health Indicator Learning (HIL) is an aspect of maintenance that learns the health conditions of equipment over time. Zhang et al. [22] proposed a data-driven approach for solving HIL problem based on model-based and model-free RL methods; Holmgren [23] presented a general-purpose maintenance planner based on Monte-Carlo tree search (MCTS); Ong et al. [24] proposed a model-free DRL algorithm, Prioritized Double Deep Q-Learning with Parameter Noise (PDDQN-PN) for predictive equipment maintenance from an equipment-based sensor network context, which can rapidly learn an optimal maintenance policy; Huang et al. [25] proposed a DDQN-based algorithm to learn the predictive maintenance policy.

Management of Engineering Systems

Decision-making for engineering systems can be formulated as an MDP or a POMDP problem [26]. Andriotis and Papakonstantinou [27] developed Deep Centralized Multi-agent Actor-Critic (DCMAC), which provides solutions for the sequential decision-making in multi-state, multi-component, partially, or fully observable stochastic engineering environments. Most studies on industrial energy management are working on modeling complex industrial processes. Huang et al. [28] developed a model-free demand response (DR) scheme for industrial facilities, with an actor-critic-based DRL algorithm to determine the optimal energy management policy.

Process Control

Automatic process control in engineering systems is to achieve a production level of consistency, economy, and safety. In contrast to the traditional design process, RL can learn appropriate closed-loop controllers by interacting with the process

and incrementally improving control behavior.

Spielberg et al. [29] proposed a DRL method for process control with the controller interacting with a process through control actions. Deep neural networks serve as function approximators to learn the control policies. In 2019, Spielberg et al. [30] also developed an adaptive model-free DRL controller for set-point tracking problems in nonlinear processes, evaluated on Single-Input-Single-Output (SISO), Multi-Input-Multi-Output (MIMO), and a nonlinear system. The results show that it can be utilized as an alternative to traditional model-based controllers.

2.2.2. Smart Grid

Smart grids are the development trend of power systems. They've been researched for years. The rise of artificial intelligence enables more complex techniques in smart grids and their future development. Zhang et al. [31] provided a review on the research and practice on DRL in smart grids, including anomaly detection, prediction, decision-making support for control, etc.

Rocchetta et al. [32] developed a DQN-based method for the optimal management of the operation and maintenance of power grids, which can exploit the information gathered from Prognostic Health Management devices, thus selecting optimal Operation and Maintenance (O&M) actions.

State estimation is critical in monitoring and managing the operation of a smart grid. An et al. [33] proposed a DQN detection (DQND) scheme to defend against data integrity attacks in AC power systems, which applies the main network and a target network to learn the detection strategy.

Wei et al. [34] proposed a recovery strategy to reclose the tripped transmission lines at the optimal time. The DDPG-based method is applied to adapt to uncertain cyber-attack scenarios and to make decisions in real-time, shown in **Figure 2**. The action in the cycle is to to reclose the tripped lines at a proper time. The reward is the transient energy including potential energy and kinetic energy.

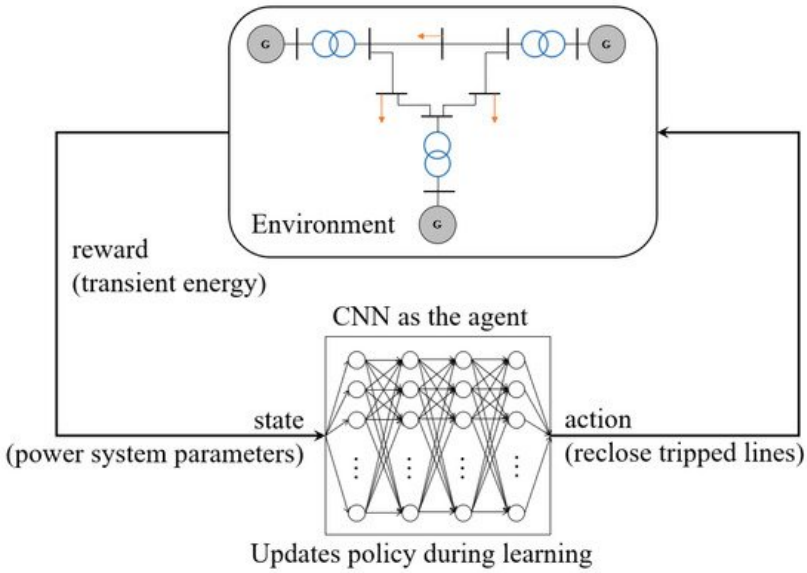

Figure 2. The schematic diagram of smart grid using DRL [34].

Mocanu et al. [35] utilized DRL in the smart grid to perform online optimization of schedules for electricity consuming devices in buildings and explored DQN and Deterministic Policy Gradient (DPG), both performing well for the minimization of the energy cost.

2.3. Communications and Networking

Modern networks, including the Internet of Things (IoT) and unmanned aerial vehicle (UAV) networks, need to make the decisions to maximize the performance under uncertainty. DRL has been applied to enable network entities to obtain optimal policies and deal with large and complex networks. Jang et al. [36] provided a survey on applications of DRL in communications and networking for traffic routing, resource sharing, and data collection. By integrating AI and blockchain, Dai et al. [37] proposed a secure and intelligent architecture for next-generation wireless networks to enable flexible and secure resource sharing and developed a caching scheme based on DRL. Also, Yang et al. [38] presented a brief review of ML applications in intelligent wireless networks.

2.4. More Topics

There are many applications based on DRL in various domains. In this section, the applications in healthcare, education, finance and aerospace will be briefly discussed.

References

1. Xiang, X.; Foo, S. Recent Advances in Deep Reinforcement Learning Applications for Solving Partially Observable Markov Decision Processes (POMDP) Problems: Part 1—Fundamentals and Applications in Games, Robotics and Natural Language Processing. Mach. Learn. Knowl. Extr. 2021, 3, 554–581.
2. Dulac-Arnold, G.; Mankowitz, D.; Hester, T. Challenges of real-world reinforcement learning. arXiv 2019, arXiv:1904.12901.
3. Finn, C.; Abbeel, P.; Levine, S. Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks. In Proceedings of the 34th International Conference on Machine Learning, Sydney, Australia, 6–11 August 2017; Volume 70, pp. 1126–1135.
4. Osband, I.; Blundell, C.; Pritzel, A.; Roy, B.V. Deep Exploration via Bootstrapped DQN. arXiv 2016, arXiv:1602.04621.
5. Mnih, V.; Kavukcuoglu, K.; Silver, D.; Rusu, A.A.; Veness, J. Human-level control through deep reinforcement learning. Nature 2015, 518, 529–533.
6. Lillicrap, T.P.; Hunt, J.J.; Alexander Pritzel, N.H.; Erez, T.; Tassa, Y.; Silver, D.; Wierstra, D. Continuous Control with Deep Reinforcement Learning. arXiv 2016, arXiv:1509.02971.
7. Dulac-Arnold, G.; Evans, R.; Sunehag, P.; Coppin, B. Reinforcement Learning in Large Discrete Action Spaces. arXiv 2015, arXiv:1512.07679.
8. Zahavy, T.; Haroush, M.; Merlis, N.; Mankowitz, D.J.; Mannor, S. Learn What Not to Learn: Action Elimination with Deep Reinforcement Learning. arXiv 2018, arXiv:1809.02121.
9. He, J.; Chen, J.; He, X.; Gao, J.; Li, L.; Deng, L.; Ostendorf, M. Deep Reinforcement Learning with an Unbounded Action Space. arXiv 2015, arXiv:1511.04636.
10. Boutilier, C.; Lu, T. Budget Allocation Using Weakly Coupled, Constrained Markov Decision Processes; ResearchGate: Berlin, Germany, 2016.
11. Bellemare, M.G.; Dabney, W.; Munos, R. A Distributional Perspective on Reinforcement Learning. In Proceedings of the International Conference on Machine Learning 2017, Sydney, Australia, 6–11 August 2017.
12. Abbeel, P.; Ng, A.Y. Apprenticeship learning via inverse reinforcement learning. In Proceedings of the Twenty-First International Conference on Machine Learning, Banff, AB, Canada, 4–8 July 2004; p. 1.
13. Hester, T.; Quinlan, M.J.; Stone, P. A Real-Time Model-Based Reinforcement Learning Architecture for Robot Control. arXiv 2011, arXiv:1105.1749.
14. Silver, D.; Huang, A.; Maddison, C.J.; Guez, A.; Sifre, L.; van den Driessche, G.; Schrittwieser, J.; Antonoglou, I.; Panneershelvam, V.; Lanctot, M.; et al. Mastering the game of Go with deep neural networks and tree search. Nature 2016, 529, 484–489.
15. Hung, C.; Lillicrap, T.P.; Abramson, J.; Wu, Y.; Mirza, M.; Carnevale, F.; Ahuja, A.; Wayne, G. Optimizing Agent Behavior over Long Time Scales by Transporting Value. arXiv 2018, arXiv:1810.06721.
16. Arjona-Medina, J.A.; Gillhofer, M.; Widrich, M.; Unterthiner, T.; Hochreiter, S. RUDDER: Return Decomposition for Delayed Rewards. arXiv 2018, arXiv:1806.07857.
17. Bazzan, A.L.; Klügl, F. Introduction to intelligent systems in traffic and transportation. Synth. Lect. Artif. Intell. Mach. Learn. 2013, 7, 1–137.
18. Haydari, A.; Yilmaz, Y. Deep Reinforcement Learning for Intelligent Transportation Systems: A Survey. arXiv 2020, arXiv:2005.00935.
19. Li, Y. Deep Reinforcement Learning: An Overview. arXiv 2018, arXiv:1701.07274.
20. Schoettler, G.; Nair, A.; Luo, J.; Bahl, S.; Ojea, J.A.; Solowjow, E.; Levine, S. Deep reinforcement learning for industrial insertion tasks with visual inputs and natural rewards. arXiv 2019, arXiv:1906.05841.
21. Li, F.; Jiang, Q.; Zhang, S.; Wei, M.; Song, R. Robot skill acquisition in assembly process using deep reinforcement learning. Neurocomputing 2019, 345, 92–102.

22. Zhang, C.; Gupta, C.; Farahat, A.; Ristovski, K.; Ghosh, D. Equipment health indicator learning using deep reinforcement learning. In Joint European Conference on Machine Learning and Knowledge Discovery in Databases; Springer: Berlin/Heidelberg, Germany, 2018; pp. 488–504.
23. Holmgren, V. General-Purpose Maintenance Planning Using Deep Reinforcement Learning and Monte Carlo Tree Search. Master's Dissertation, Linköping University, Linköping, Sweden, December 2019.
24. Ong, K.S.H.; Niyato, D.; Yuen, C. Predictive Maintenance for Edge-Based Sensor Networks: A Deep Reinforcement Learning Approach. In Proceedings of the 2020 IEEE 6th World Forum on Internet of Things (WF-IoT), New Orleans, LA, USA, 2–16 June 2020; pp. 1–6.
25. Huang, J.; Chang, Q.; Arinez, J. Deep reinforcement learning based preventive maintenance policy for serial production lines. Expert Syst. Appl. 2020, 160, 113701.
26. Andriotis, C.; Papakonstantinou, K. Life-cycle policies for large engineering systems under complete and partial observability. In Proceedings of the 13th International Conference on Applications of Statistics and Probability in Civil Engineering (ICASP13), Seoul, Korea, 26–30 May 2019.
27. Andriotis, C.P.; Papakonstantinou, K.G. Managing engineering systems with large state and action spaces through deep reinforcement learning. arXiv 2018, arXiv:1811.02052.
28. Huang, X.; Hong, S.H.; Yu, M.; Ding, Y.; Jiang, J. Demand Response Management for Industrial Facilities: A Deep Reinforcement Learning Approach. IEEE Access 2019, 7, 82194–82205.
29. Spielberg, S.P.K.; Gopaluni, R.B.; Loewen, P.D. Deep reinforcement learning approaches for process control. In Proceedings of the 2017 6th International Symposium on Advanced Control of Industrial Processes (AdCONIP), Taipei, Taiwan, 28–31 May 2017; pp. 201–206.
30. Spielberg, S.; Tulsyan, A.; Lawrence, N.P.; Loewen, P.D.; Bhushan Gopaluni, R. Toward self-driving processes: A deep reinforcement learning approach to control. AIChE J. 2019, 65, e16689.
31. Zhang, D.; Han, X.; Deng, C. Review on the research and practice of deep learning and reinforcement learning in smart grids. CSEE J. Power Energy Syst. 2018, 4, 362–370.
32. Rocchetta, R.; Bellani, L.; Compare, M.; Zio, E.; Patelli, E. A reinforcement learning framework for optimal operation and maintenance of power grids. Appl. Energy 2019, 241, 291–301.
33. An, D.; Yang, Q.; Liu, W.; Zhang, Y. Defending against data integrity attacks in smart grid: A deep reinforcement learning-based approach. IEEE Access 2019, 7, 110835–110845.
34. Wei, F.; Wan, Z.; He, H. Cyber-Attack Recovery Strategy for Smart Grid Based on Deep Reinforcement Learning. IEEE Trans. Smart Grid 2020, 11, 2476–2486.
35. Mocanu, E.; Mocanu, D.C.; Nguyen, P.H.; Liotta, A.; Webber, M.E.; Gibescu, M.; Slootweg, J.G. On-Line Building Energy Optimization Using Deep Reinforcement Learning. IEEE Trans. Smart Grid 2019, 10, 3698–3708.
36. Jang, K.; Vinitsky, E.; Chalaki, B.; Remer, B.; Beaver, L.; Malikopoulos, A.A.; Bayen, A. Simulation to scaled city: Zero-shot policy transfer for traffic control via autonomous vehicles. In Proceedings of the 10th ACM/IEEE International Conference on Cyber-Physical Systems, Montreal, QC, Canada, 16–18 April 2019; pp. 291–300.
37. Dai, Y.; Xu, D.; Maharjan, S.; Chen, Z.; He, Q.; Zhang, Y. Blockchain and deep reinforcement learning empowered intelligent 5G beyond. IEEE Netw. 2019, 33, 10–17.
38. Yang, H.; Xie, X.; Kadoch, M. Machine Learning Techniques and A Case Study for Intelligent Wireless Networks. IEEE Netw. 2020, 34, 208–215.
