machine learning & knowledge extraction

MDPI

Review

Recent Advances in Deep Reinforcement Learning Applications for Solving Partially Observable Markov Decision Processes (POMDP) Problems Part 2—Applications in Transportation, Industries, Communications and Networking and More Topics

Xuanchen Xiang *, Simon Foo * and Huanyu Zang *

Department of Electrical and Computer Engineering, FAMU-FSU College of Engineering, Tallahassee, FL 32310, USA

* Correspondence: xx16@my.fsu.edu (X.X.); foo@eng.famu.fsu.edu (S.F.); hz16b@my.fsu.edu (H.Z.)

Abstract: The two-part series of papers provides a survey on recent advances in Deep Reinforcement Learning (DRL) for solving partially observable Markov decision processes (POMDP) problems. 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. The fact that the agent has limited access to the information of the environment enables AI to be applied efficiently in most fields that require self-learning. It's essential to have an organized investigation—we can make good comparisons and choose the best structures or algorithms when applying DRL in various applications. The first part of the overview introduces Markov Decision Processes (MDP) problems and Reinforcement Learning and applications of DRL for solving POMDP problems in games, robotics, and natural language processing. In part two, we continue to introduce applications in transportation, industries, communications and networking, etc. and discuss the limitations of DRL.

Keywords: reinforcement learning; deep reinforcement learning; Markov decision process; partially observable markov decision process

Citation: Xiang, X.; Foo, S.; Zang, H. Recent Advances in Deep Reinforcement Learning Applications for Solving Partially Observable Markov Decision Processes (POMDP) Problems Part 2—Applications in Transportation, Industries, Communications and Networking and More Topics. *Mach. Learn. Knowl. Extr.* **2021**, *3*, 863–878. https://doi.org/10.3390/make3040043

Academic Editor: Andreas Holzinger

Received: 23 September 2021
Accepted: 23 October 2021
Published: 28 October 2021

1. Applications

1.1. Transportation

An intelligent transportation system (ITS) [1] 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 [2] presented a comprehensive survey on DRL for ITS.

1.1.1. Traffic Signal Control (TSC)

An application of ML in transportation is the scheduling of traffic signals in multi-intersection vehicular networks. Automatic signal control makes traffic flow more efficient and reasonable and alleviates traffic congestion.

Arel et al. [3] explained the configuration of RL where the states are based on traffic statistics, with each element being the traffic flow at each lane. The agent selects an action based on vehicle positions, according to its policy. Haydari and Yilmaz [2] reviewed two popular types of state representations in intersections. The first format is an image-like representation called Discrete Traffic State Encoding (DTSE). It acquires high resolution and practical information from the intersection. Four characteristics, including speed and

position of vehicles, signal phases, and accelerations, are selected in different research, shown in separate arrays in DTSE. The second approach is forming a feature-based vector. The average or total value of information for each lane is represented on a vector instead of using vehicle-based state representation. The features include queue length, cumulative waiting time, the average speed on a lane, phase duration, and the number of vehicles in each lane. The two representations are shown in Figure 1. State and reward settings are also discussed in [2].

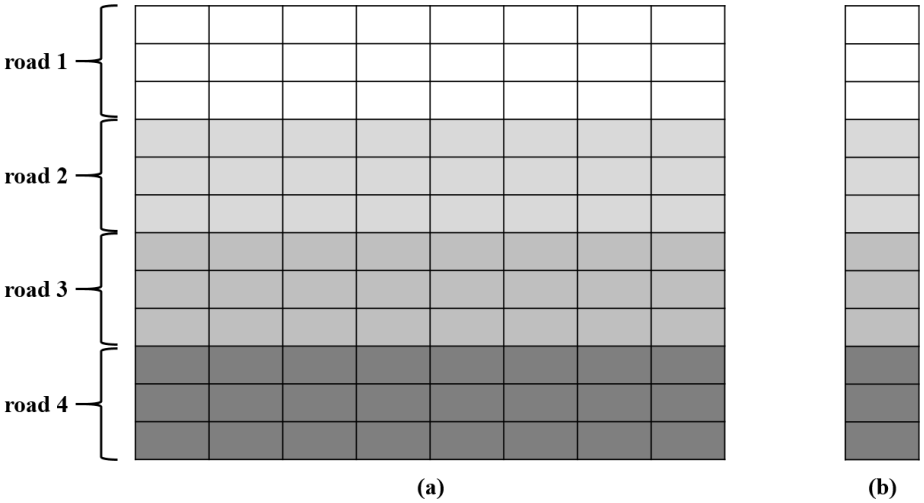

Figure 1. Two popular types of state representation in an intersection with four roads (in four different colors) and three lanes in each road: (**a**) DTSE matrix—Each cell represents one vehicle; (**b**) Feature-based state vector – Each cell represents a lane [2].

Li [4] discussed works regarding adaptive control for signal control: MARLIN-ATSC proposed by El-Tantawy et al. [5]; Van der Pol and Oliehoek [6] combined Deep Q Networks (DQN) [7] and coordination algorithm; Mannion et al. [8] provided an experimental review of DRL for adaptive traffic signal control. For more surveys, see [2,9–13].

Genders and Razavi [14] proposed the discrete traffic state encoding, which is information-dense, as the input to the DQN networks for traffic signal control agent (DQTSCA) and evaluated state representations from low to high-resolution using Asynchronous Advantage Actor Critic (A3C) in [15]. Garg et al. [16] built a traffic simulator on a 3D virtual reality software, Unity3d, taking collision count, speed of vehicles across the intersections, dynamic generation, etc. into consideration, and created a simulation environment closely based on the real-world traffic specifications. Rodrigues and Azevedo [17] developed an open-source callback-based framework (CAREL) integrated with AIMSUN, for testing as a benchmark. Wei et al. [18] proposed a decentralized RL method for multi-intersection traffic signal control on arterial traffic, with each intersection with an individual control agent. Wang et al. [19] introduced a Double Dueling Deep Q Network (3DQN) with high-resolution event-based data, which is collected directly from vehicle-actuated detectors.

Recently, Ma and Wu [20] proposed Feudal Multi-agent Advantage Actor-Critic (FMA2C), an extension of MA2C [21] with feudal hierarchy, with each split region controlled by an agent. Wu et al. [22] presented multi-agent recurrent deep deterministic policy gradient (MARDDPG), based on Deep Deterministic Policy Gradient (DDPG) [23]. Xu et al. [24] used a data-driven approach to find critical nodes, which can cause a reduction in traffic efficiency. They then introduced a policy gradient method on these nodes. This method can effectively lower the average delay and travel time.

In 2020, Haydari and Yilmaz [2] provided tables of outlines of single and multiple agent RL approaches for Traffic Signal Control (TSC), DRL methods for TSC, and DRL solutions for other ITS applications.

1.1.2. Autonomous Driving

Autonomous driving is an essential topic of ITS. TORCS is often used as the auto-driving simulator for algorithms such as DDPG [23], Deep Deterministic Actor Critic Algorithm (DDAC) [25], Fine Grained Action Repetition (FiGAR) [26], Normalized Actor-Critic (NAC) [27], etc., as mentioned in [28]. Due to safety concerns, road test for algorithms on auto-vehicles hasn't been widely applied. But researchers built many open-source simulators, and most methods were evaluated in simulation. Kang et al. [29] provided an overview of driving datasets and virtual testing environments. There are several tasks involved in autonomous driving, including motion planning, overtaking, merging, lane change, auto-parking, etc., see [30,31] for surveys of DRL algorithms in auto driving.

Sim-to-Real

Osiński et al. [32] presented simulation-Based RL for real-world autonomous driving. They used RL in simulation to make the driving system control the real-world vehicle and achieved sim-to-real policy transfer.

Navigation

Navigation is a fundamental task in autonomous driving, and DRL has been proven to be effective in navigation problems: Fayjie et al. [33] presented a DQN-based approach for navigation in the urban environment, and Isele et al. [34] used a DQN-based method for navigating in occluded intersections.

Pusse and Klusch [35] introduced a hybrid solution, HyLEAP, for pedestrian collision-free navigation. It combines selected POMDP planning methods and DRL prior to other individual methods regarding German In-Depth Accident Study (GIDAS) pedestrian safety.

Lane Change

Sharifzadeh et al. [36] proposed an inverse reinforcement learning (IRL) approach with DQN to extract the rewards for collision-free lane changing. The agent can perform human-like lane changing behavior; Hoel et al. [37] used a Double DQN agent for speed change and lane change, and overtaking cases, which outperforms the combination of the Intelligent Driver Model (IDM) and Minimizing Overall Braking Induced by Lane changes (MOBIL) model in highway driving; Shi et al. [38] applied Hierarchical DQN: Firstly, DQN is used to decide when to perform the lane change. Secondly, a Q-function in a quadratic form is designed for car-following and the gap in the target lane. Lastly, the execution step is to perform movement. Wang et al. [39] later presented a rule-based DQN, which outperforms individual DQN or rule-based methods. Regarding the driver assistant systems, Min et al. [40] proposed a supervisor agent using Quantile Regression Deep Q Network (QR-DQN) for lane changing and other control. Ye et al. [41] adopted DDPG for the training and high-fidelity virtual simulation environment VISSIM, getting better results than IDM and Constant Time Headway (CTH).

Decision Making (and Optimum Control)

Making safe and effective decisions in complex traffic environments is crucial in auto-driving. The method must be general to handle the changing situations. The problems include when to change lanes, or whether or not to stop at an intersection, etc.

Due to the varying environments, the uncertainty should be considered when applying autonomous driving to the real world. Modeling the problem as a POMDP will be necessary. The requirement for storing observations can cause inefficiency. Qiao et al. [42] introduced Hierarchical Options MDP (HOMDP), which learns discrete options in the high-level process and low-level continuous actions simultaneously.

Combining the concepts of planning and learning, Hoel et al. [43] introduced Monte Carlo tree search (MCTS) and DRL framework for tactical decision making, based on the AlphaGo Zero algorithm with a continuous state space. The agents outperform baselines.

Autonomous driving is a typical multi-agent setting, Yu et al. [44] employed coordination-graph-based multi-agent RL (MARL) approaches to achieve coordinated maneuvers for multiple vehicles. The method can achieve a high level of safety by properly coordinating vehicles' overtaking maneuvers.

With the concept of Connected and Automated Vehicles (CAVs), a vehicle's behaviors are based on shared information. Zhou et al. [45] proposed a DDPG-based car-following model and trained CAVs to obtain appropriate behaviors to improve travel efficiency, fuel consumption, and safety at signalized intersections in real-time.

Path Planning

Makantasis et al. [46] considered path planning for an autonomous vehicle that moves on a freeway. The experiments show that the DDQN-derived driving policy can achieve better performance comparing to DP (Dynamic Programming) or SUMO policies; Qian et al. [47] proposed a planning features-based deep behavior decision method (PFBD), trained with Twin Delayed DDPG (TD3), to select an optimal maneuver.

Pedestrian Detection

For pedestrian detection, Chae et al. [48] proposed an autonomous braking system using DQN.

1.1.3. Other Applications in ITS

Ramp Metering

Belletti et al. [49] presented Multi-Task DRL for control of systems modeled by discretized non-linear Partial Differential Equations (PDEs) and achieved expert-level control of Ramp Metering. Chalaki et al. [50] developed a zero-shot transfer of a policy from simulation to the University of Delaware's Scaled Smart City (UDSSC) testbed. The adversarial multi-agent policy improves system efficiency even under stochastic. Based on this, Jang et al. [51] trained two policies, and the noised policy significantly outperformed the noise-free version.

Energy Management

Qi et al. [52] designed a DQN-based PHEV (plug-in hybrid electric vehicles) energy management system to autonomously splits fuel/electricity from interactions between the car and the environment, making the model capable of achieving energy savings.

1.2. Industrial Applications

1.2.1. 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 [4].

Robotics, including manipulation, locomotion, etc., will prevail in all aspects of industrial applications, which was mentioned in [28]. For example, Schoettler et al. [53] discussed insertion tasks, particularly in industrial applications; Li et al. [54] 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. [55] proposed a data-driven approach for solving HIL problem based on model-based and model-free RL methods; Holmgren [56] presented a general-purpose maintenance planner based on Monte-Carlo tree search (MCTS); Ong et al. [57] 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. [58] 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 [59]. Andriotis and Papakonstantinou [60] 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. [61] 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. [62] 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. [63] 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.

1.2.2. 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. [64] 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. [65] 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. [66] 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. [67] 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.

Mocanu et al. [68] 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.

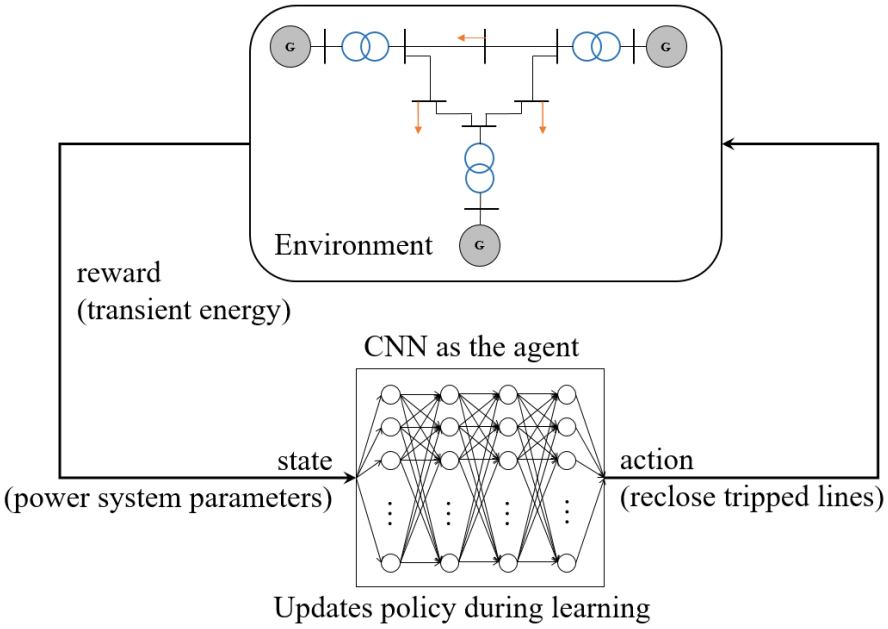

Figure 2. The schematic diagram of smart grid using DRL [67].

1.3. 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. [51] 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. [69] 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. [70] presented a brief review of ML applications in intelligent wireless networks.

1.3.1. Internet of Things (IoT)

The Internet of Things (IoT) connects a great amount of devices to the Internet, where the devices collect and share sensory data to reflect the status of the physical world. Autonomous IoT (AIoT) integrates IoT, ML, and autonomous control. AI is a promising method to achieve autonomy, for decision making. Lei et al. [71] proposed a general 3-layer model for the applications of RL/DRL in AIoT. For each layer, the state is the system state and the reward is the performance of the system. The action of the loop is the control to the layer systems, as shown in Figure 3.

Industrial Internet of Things (IIoT)

Blockchain is a promising solution for data storing/processing/sharing securely and efficiently in the industrial Internet of things (IIoT). Blockchain-enabled IIoT systems can utilize DRL techniques to improve the performance [72,73].

Mobile Edge Computing (MEC)

Mobile Edge Computing (MEC) is a promising technology to extend the services to the edge of the IoT system, and DRL has been successfully applied in the MEC networks in recent years [74–76]. Zhu et al. [77] discussed DRL in caching transient data. Chen et al. [78] proposed intelligent resource allocation framework (iRAF) to solve the resource allocation problem for collaborative mobile edge computing (CoMEC). The technology of fog computing is a promising paradigm for IoT to provide proximity services. There are research focusing on utilizing DRL in fog-enabled IoT [79–81].

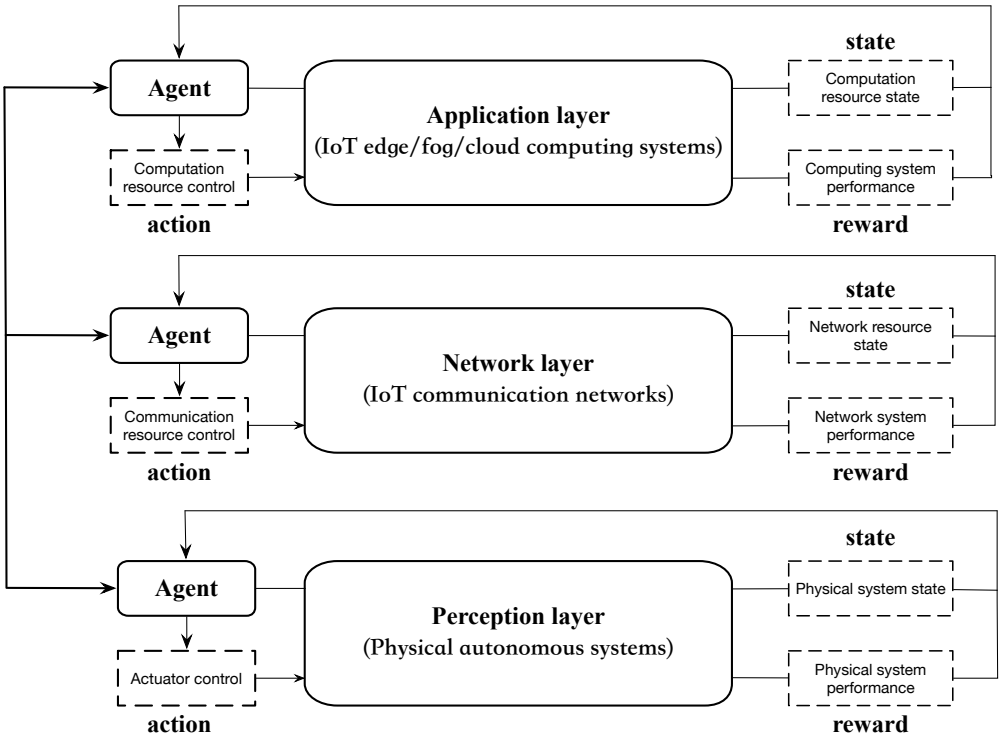

Figure 3. General RL/DRL model for autonomous IoT [71].

Others

Zhu et al. [82] proposed a DQN-based transmission scheduling mechanism for the cognitive radio-based IoT (CIoT) to find the optimal strategy to transmit packets of different buffers among multiple channels to maximize the system utility. Ferdowsi and Saad [83] proposed a DRL-based watermarking algorithm for dynamic authentication of IoT signals to detect cyber-attacks. Jay et al. [84] presented Aurora, a congestion control protocol powered by DRL.

1.3.2. Connected Vehicles

Vehicle networks share information and entertainment to drivers and vehicles by connecting services, content, and application providers through wireless networks, including safety warnings, managing, playing audio, navigation, delivering entertainment, social networking, making phone calls, autonomous driving, etc. This section is similar to what's discussed in Section 1.1 but will focus more on communications between vehicles.

He et al. [85] proposed an integrated framework that enables dynamic orchestration of networking, caching, and computing, to improve the performance of next-generation vehicular networks. Doddalinganavar et al. [86] provided a survey on DRL protocols in Vehicular Adhoc Networks (VANETS).

Computing and Caching

DRL has also been utilized successfully in computing and caching in-vehicle networks. Tan and Hu [87] developed a DRL-based multi-timescale framework, with mobility-aware reward estimation. Ning et al. [88] and Liu et al. [89] provided DRL insights for vehicular edge computing and constructing an intelligent offloading system. Ning et al. [90] developed an intent-based traffic control system based on DRL for the 5G-envisioned Internet of Vehicles (IoCVs).

Resource Allocation

In Vehicle-to-Vehicle (V2V) networks, device-to-device (D2D) communications provide direct local message dissemination with low latency and energy consumption. Effec-

tive resource allocation mechanisms are necessary to manage the interference between the D2D links and the cellular links. Ye et al. [91] presented a decentralized resource allocation mechanism for V2V communications based on DRL that can be employed in both unicast and broadcast scenarios.

Traffic Scheduling

Chinchali et al. [92] presented a DRL-based scheduler that can adapt to highly dynamic traffic and various reward functions set by network operators to optimally schedule traffic, named HVFT-RL (High Volume Flexible Time).

Others

The vehicle-to-infrastructure (V2I) communication via millimeter-wave (mmWave) base stations is crucial for the operation of 5G ITS, which offers high capacity channel resources toward connected vehicles. There exists the cell association and resource allocation problem called CARA. Kwon and Kim [93] proposed the 3-tier heterogeneous vehicular network (HetVNet) using a multi-agent deep deterministic policy gradient (MADDPG) approach to solve CARA problems.

With the vehicle sensors, it's still tricky to detect objects occluded by other moving/static obstacles. [94] presented a cooperative perception scheme with DRL to select data to transmit, thus enhancing the detection accuracy. The Cooperative & Intelligent Vehicle Simulation (CIVS) Platform was developed to evaluate the scheme, and the results show significant improvement accuracy compared to the baseline.

1.3.3. Resources Management

Resources management problems are vital in all applications, as discussed in previous sections. In systems and networking, they often appear as online decision-making tasks where solutions depend on understanding the workload and environment. In these situations, DRL is helpful to deal with resource management problems. Mao et al. [95] presented DeepRM in 2016, which translates the problem with multiple resource demands into a learning problem. Li et al. [96] considered resources management in network slicing, and Zhang et al. [97] introduced intelligent cloud resource management with DRL.

1.4. 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.

1.4.1. Healthcare

RL is a suitable method to develop robust solutions for many healthcare domains, where a long and sequential procedure usually characterizes diagnosing decisions or treatment regimes.

Esteva et al. [98] presented a guide to DL in healthcare, in which DRL is discussed in the context of robotic-assisted surgery. Liu et al. [99] presented a survey of DRL methods on clinical decision support.

Yu et al. [100] discussed the broad applications of RL in healthcare, including dynamic treatment regimes in chronic diseases and critical care, automated medical diagnosis, and other control or scheduling domains in many aspects of a healthcare system. As shown in Figure 4, RL is mainly used in dynamic treatment regimes (DTRs) [101], automated medical diagnosis and others.

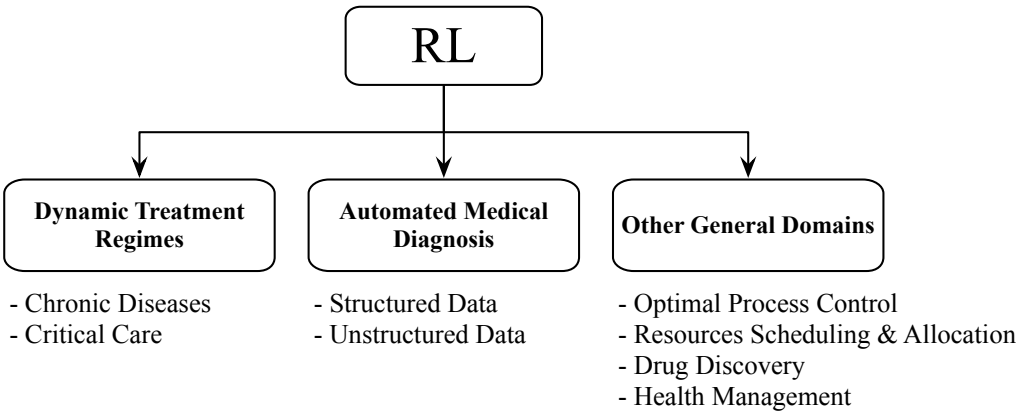

Figure 4. The outline of applications of RL in healthcare [100].

1.4.2. Education

Reddy et al. [102] mentioned that guiding students with a sequence of lessons and helping them retain knowledge is one of the central challenges in education. They studied a DRL method to learn flexible and scalable teaching policies that select the following item to review.

Zheng et al. [103] proposed a DQN-based framework for online personalized news recommendation.

1.4.3. Finance

DRL can provide a long-term strategy to maximize cumulative reward, making it popular in trading, stocking, and marketing. Like AlphaZero and other games, there are similar rules in trading. The most widely used mechanism in financial markets is the "continuous double auction order book with time priority", based on which, Ritter [104] discussed RL algorithms mathematically. See [105] for a survey on RL in economics and finance.

Financial trading is crucial to investment companies, and DRL has developed as a good option to generalize trading strategies or to analyze [106,107]. Wang et al. [108] proposed AlphaStock for better quantitative trading (QT) strategies.

Financial portfolio management is the procedure of constant redistribution of a fund into different financial products. DRL has been proved to be effective for financial portfolio management [109,110].

1.4.4. Aerospace

Increasingly complex space missions have encouraged the development of autonomous command and control approaches for handling high-dimensional, continuous observations and action spaces with hard-to-analyze behavior. DRL techniques have been researched for providing safety and performance in aerospace. Harris and Schaub [111] and Harris et al. [112] have examined DRL methods in spacecraft command and control and decision making.

DRL has also been researched in cognitive aerospace communications [113], UAV networks [114] and monitoring tasks [115].

2. Discussions

2.1. Deep Reinforcement Learning Limitations

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. [28]. 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. [116] addressed nine significant challenges of practical RL in the real world. They presented examples for each challenge and provided some references for deploying RL:

1. Modeling the real world is complex. Many systems cannot be directly trained on. An off-line off-policy approach [116] 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.
2. 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. [117] proposed Model Agnostic Meta-Learning (MAML) to learn within a distribution with few shot learning. Osband et al. [118] 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 [7] and DDPG [23].
3. Real-world environments usually have massive and continuous state and action spaces. Dulac-Arnold et al. [119] addressed the challenge for sizeable discrete action spaces. Action-Elimination Deep Q-Network (AE-DQN) [120] and Deep Reinforcement Relevance Network (DRRN) [121] also deals with the issue.
4. The learned policy might violate the safety constraints. Constrained MDP (CMDP) [116] and budgeted MDP [122] take the constraint components into consideration during training.
5. Considering POMDP problems, Dulac-Arnold et al. [116] presented Robust MDPs, where the learned policy maximizes the worst-case value function.
6. Formulating multi-dimensional reward functions is usually necessary and complicated. Distributional DQN Bellemare et al. [123] models the percentile distribution of the rewards. Dulac-Arnold et al. [116] presented multi-objective analysis and formulated the global reward function as a linear combination of sub-rewards. Abbeel and Ng [124] gave an algorithm is based on inverse RL to try to recover the unknown reward function.
7. Policy explainability is vital for real-world policies as humans operate the systems.
8. Policy inference should be made in real-time at the control frequency of the system. Hester et al. [125] presented a parallel real-time architecture for model-based RL. AlphaGo [126] improves with more rollouts rather than running at a specific frequency.
9. Most natural systems have delays in the perception of the states, the actuators, or the return. Hung et al. [127] proposed a memory-based algorithm where agents use recall of memories to credit actions from the past. Arjona-Medina et al. [128] introduced RUDDER (Return Decomposition for Delayed Rewards) to learn long-term credit assignments for delayed rewards.

2.2. Summary

This is the second part of the two-part series of survey papers. In the first part [28], the fundamental concepts and some applications have been discussed. In this part, we continue presented applications in more domains, where DRL hasn't been as widely employed as in Gaming and Robotics.

In an intelligent transportation system, we discussed Traffic Signal Control (TSC), auto-driving, and more control applications. However, testing the policies in the real world is a huge barrier. In industries, communications and networking applications, DRL is proven to be an alternative to conventional methods. Note that there are some inevitable overlapping among those applications. For example, applying DRL in the game is also part of the application in transportation; robotics can also be in industrial or healthcare applications.

As we witnessed, many research groups pushed forward the fast development of DRL, including OpenAI, DeepMind, AI research office in Alberta, and research center led by Rich Sutton, and more. Many new algorithms are being developed and discussed. This overview provides an organized investigation to help us get more familiar with the usability of different methods and architectures. And it's always inspiring to optimize the strategies and to overcome the limitations.

Author Contributions: Conceptualization, X.X. and S.F.; methodology, X.X.; formal analysis, X.X.; investigation, X.X., S.F. and H.Z.; resources, X.X., S.F. and H.Z.; writing—original draft preparation, X.X.; writing—review and editing, X.X., S.F. and H.Z.; visualization, X.X.; supervision, X.X. and S.F.; funding acquisition, S.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Acknowledgments: The authors would like to express their appreciation to friends and colleagues who had provided assistance during the preparation of this paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bazzan, A.L.; Klügl, F. Introduction to intelligent systems in traffic and transportation. *Synth. Lect. Artif. Intell. Mach. Learn.* **2013**, *7*, 1–137. [CrossRef]
2. Haydari, A.; Yilmaz, Y. Deep Reinforcement Learning for Intelligent Transportation Systems: A Survey. *arXiv* **2020**, arXiv:2005.00935.
3. Arel, I.; Liu, C.; Urbanik, T.; Kohls, A.G. Reinforcement learning-based multi-agent system for network traffic signal control. *IET Intell. Transp. Syst.* **2010**, *4*, 128–135. [CrossRef]
4. Li, Y. Deep Reinforcement Learning: An Overview. *arXiv* **2018**, arXiv:1701.07274.
5. El-Tantawy, S.; Abdulhai, B.; Abdelgawad, H. Multiagent Reinforcement Learning for Integrated Network of Adaptive Traffic Signal Controllers (MARLIN-ATSC): Methodology and Large-Scale Application on Downtown Toronto. *IEEE Trans. Intell. Transp. Syst.* **2013**, *14*, 1140–1150. [CrossRef]
6. Van der Pol, E.; Oliehoek, F.A. Coordinated deep reinforcement learners for traffic light control. In Proceedings of the Learning, Inference and Control of Multi-Agent Systems, Barcelona, Spain, 5–10 December 2016.
7. Mnih, V.; Kavukcuoglu, K.; Silver, D.; Rusu, A.A.; Veness, J. Human-level control through deep reinforcement learning. *Nature* **2015**, *518*, 529–533. [CrossRef]
8. Mannion, P.; Duggan, J.; Howley, E. An experimental review of reinforcement learning algorithms for adaptive traffic signal control. In *Autonomic Road Transport Support Systems*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 47–66.
9. Gregurić, M.; Vujić, M.; Alexopoulos, C.; Miletić, M. Application of Deep Reinforcement Learning in Traffic Signal Control: An Overview and Impact of Open Traffic Data. *Appl. Sci.* **2020**, *10*, 4011. [CrossRef]
10. Muresan, M.; Fu, L.; Pan, G. Adaptive traffic signal control with deep reinforcement learning an exploratory investigation. *arXiv* **2019**, arXiv:1901.00960.
11. Gong, Y. Improving Traffic Safety and Efficiency by Adaptive Signal Control Systems Based on Deep Reinforcement Learning. Ph.D. Thesis, University of Central Florida, Orlando, FL, USA, 2020.
12. Tan, K.L.; Poddar, S.; Sarkar, S.; Sharma, A. Deep Reinforcement Learning for Adaptive Traffic Signal Control. In Proceedings of the Dynamic Systems and Control Conference, Park City, UT, USA, 8–11 October 2019; Volume 59162, p. V003T18A006.
13. Guo, J. Decentralized Deep Reinforcement Learning for Network Level Traffic Signal Control. *arXiv* **2020**, arXiv:2007.03433.
14. Genders, W.; Razavi, S. Using a deep reinforcement learning agent for traffic signal control. *arXiv* **2016**, arXiv:1611.01142.

15. Genders, W.; Razavi, S. Evaluating reinforcement learning state representations for adaptive traffic signal control. *Procedia Comput. Sci.* **2018**, *130*, 26–33. [CrossRef]
16. Garg, D.; Chli, M.; Vogiatzis, G. Deep Reinforcement Learning for Autonomous Traffic Light Control. In Proceedings of the 2018 3rd IEEE International Conference on Intelligent Transportation Engineering (ICITE), Singapore, 3–5 September 2018; pp. 214–218.
17. Rodrigues, F.; Azevedo, C.L. Towards Robust Deep Reinforcement Learning for Traffic Signal Control: Demand Surges, Incidents and Sensor Failures. In Proceedings of the 2019 IEEE Intelligent Transportation Systems Conference (ITSC), Auckland, New Zealand, 27–30 October 2019; pp. 3559–3566.
18. Wei, H.; Chen, C.; Wu, K.; Zheng, G.; Yu, Z.; Gayah, V.; Li, Z. Deep Reinforcement Learning for Traffic Signal Control along Arterials. In Proceedings of the 2019, DRL4KDD '19, Anchorage, AK, USA, 5 August 2019.
19. Wang, S.; Xie, X.; Huang, K.; Zeng, J.; Cai, Z. Deep reinforcement learning-based traffic signal control using high-resolution event-based data. *Entropy* **2019**, *21*, 744. [CrossRef] [PubMed]
20. Ma, J.; Wu, F. Feudal Multi-Agent Deep Reinforcement Learning for Traffic Signal Control. In Proceedings of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS), Auckland, New Zealand, 9–13 May 2020.
21. Chu, T.; Wang, J.; Codecà, L.; Li, Z. Multi-Agent Deep Reinforcement Learning for Large-Scale Traffic Signal Control. *IEEE Trans. Intell. Transp. Syst.* **2020**, *21*, 1086–1095. [CrossRef]
22. Wu, T.; Zhou, P.; Liu, K.; Yuan, Y.; Wang, X.; Huang, H.; Wu, D.O. Multi-Agent Deep Reinforcement Learning for Urban Traffic Light Control in Vehicular Networks. *IEEE Trans. Veh. Technol.* **2020**, *69*, 8243–8256. [CrossRef]
23. 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.
24. Xu, M.; Wu, J.; Huang, L.; Zhou, R.; Wang, T.; Hu, D. Network-wide traffic signal control based on the discovery of critical nodes and deep reinforcement learning. *J. Intell. Transp. Syst.* **2020**, *24*, 1–10. [CrossRef]
25. Sallab, A.E.; Abdou, M.; Perot, E.; Yogamani, S. End-to-End Deep Reinforcement Learning for Lane Keeping Assist. *arXiv* **2016**, arXiv:1612.04340.
26. Sharma, S.; Lakshminarayanan, A.S.; Ravindran, B. Learning To Repeat: Fine Grained Action Repetition For Deep Reinforcement Learning. *arXiv* **2017**, arXiv:1702.06054.
27. Gao, Y.; Xu, H.; Lin, J.; Yu, F.; Levine, S.; Darrell, T. Reinforcement Learning from Imperfect Demonstrations. *arXiv* **2018**, arXiv:1802.05313.
28. 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. [CrossRef]
29. Kang, Y.; Yin, H.; Berger, C. Test Your Self-Driving Algorithm: An Overview of Publicly Available Driving Datasets and Virtual Testing Environments. *IEEE Trans. Intell. Veh.* **2019**, *4*, 171–185. [CrossRef]
30. Kiran, B.R.; Sobh, I.; Talpaert, V.; Mannion, P.; Sallab, A.A.A.; Yogamani, S.; Pérez, P. Deep reinforcement learning for autonomous driving: A survey. *arXiv* **2020**, arXiv:2002.00444.
31. Grigorescu, S.; Trasnea, B.; Cocias, T.; Macesanu, G. A survey of deep learning techniques for autonomous driving. *J. Field Robot.* **2020**, *37*, 362–386. [CrossRef]
32. Osiński, B.; Jakubowski, A.; Miłoś, P.; Zięcina, P.; Galias, C.; Homoceanu, S.; Michalewski, H. Simulation-based reinforcement learning for real-world autonomous driving. *arXiv* **2020**, arXiv:1911.12905.
33. Fayjie, A.R.; Hossain, S.; Oualid, D.; Lee, D.J. Driverless car: Autonomous driving using deep reinforcement learning in urban environment. In Proceedings of the 2018 15th International Conference on Ubiquitous Robots (UR), Honolulu, HI, USA, 26–30 June 2018; pp. 896–901.
34. Isele, D.; Rahimi, R.; Cosgun, A.; Subramanian, K.; Fujimura, K. Navigating occluded intersections with autonomous vehicles using deep reinforcement learning. In Proceedings of the 2018 IEEE International Conference on Robotics and Automation (ICRA), Brisbane, Australia, 21–25 May 2018; pp. 2034–2039.
35. Pusse, F.; Klusch, M. Hybrid Online POMDP Planning and Deep Reinforcement Learning for Safer Self-Driving Cars. In Proceedings of the 2019 IEEE Intelligent Vehicles Symposium (IV), Paris, France, 9–12 June 2019; pp. 1013–1020.
36. Sharifzadeh, S.; Chiotellis, I.; Triebel, R.; Cremers, D. Learning to Drive using Inverse Reinforcement Learning and Deep Q-Networks. *arXiv* **2016**, arXiv:1612.03653.
37. Hoel, C.J.; Wolff, K.; Laine, L. Automated speed and lane change decision making using deep reinforcement learning. In Proceedings of the 2018 21st International Conference on Intelligent Transportation Systems (ITSC), Maui, HI, USA, 4–7 November 2018; pp. 2148–2155.
38. Shi, T.; Wang, P.; Cheng, X.; Chan, C.Y.; Huang, D. Driving Decision and Control for Autonomous Lane Change based on Deep Reinforcement Learning. *arXiv* **2019**, arXiv:1904.10171.
39. Wang, J.; Zhang, Q.; Zhao, D.; Chen, Y. Lane change decision-making through deep reinforcement learning with rule-based constraints. In Proceedings of the 2019 International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, 14–19 July 2019; pp. 1–6.
40. Min, K.; Kim, H.; Huh, K. Deep Distributional Reinforcement Learning Based High-Level Driving Policy Determination. *IEEE Trans. Intell. Veh.* **2019**, *4*, 416–424. [CrossRef]

41. Ye, Y.; Zhang, X.; Sun, J. Automated vehicle's behavior decision making using deep reinforcement learning and high-fidelity simulation environment. *Transp. Res. Part Emerg. Technol.* **2019**, *107*, 155–170. [CrossRef]
42. Qiao, Z.; Muelling, K.; Dolan, J.; Palanisamy, P.; Mudalige, P. Pomdp and hierarchical options mdp with continuous actions for autonomous driving at intersections. In Proceedings of the 2018 21st International Conference on Intelligent Transportation Systems (ITSC), Maui, HI, USA, 4–7 November 2018; pp. 2377–2382.
43. Hoel, C.; Driggs-Campbell, K.R.; Wolff, K.; Laine, L.; Kochenderfer, M.J. Combining Planning and Deep Reinforcement Learning in Tactical Decision Making for Autonomous Driving. *arXiv* **2019**, arXiv:1905.02680.
44. Yu, C.; Wang, X.; Xu, X.; Zhang, M.; Ge, H.; Ren, J.; Sun, L.; Chen, B.; Tan, G. Distributed multiagent coordinated learning for autonomous driving in highways based on dynamic coordination graphs. *IEEE Trans. Intell. Transp. Syst.* **2019**, *21*, 735–748. [CrossRef]
45. Zhou, M.; Yu, Y.; Qu, X. Development of an efficient driving strategy for connected and automated vehicles at signalized intersections: A reinforcement learning approach. *IEEE Trans. Intell. Transp. Syst.* **2019**, *21*, 433–443. [CrossRef]
46. Makantasis, K.; Kontorinaki, M.; Nikolos, I.K. A Deep Reinforcement-Learning-based Driving Policy for Autonomous Road Vehicles. *arXiv* **2019**, arXiv:1907.05246.
47. Qian, L.; Xu, X.; Zeng, Y.; Huang, J. Deep, Consistent Behavioral Decision Making with Planning Features for Autonomous Vehicles. *Electronics* **2019**, *8*, 1492. [CrossRef]
48. Chae, H.; Kang, C.M.; Kim, B.; Kim, J.; Chung, C.C.; Choi, J.W. Autonomous braking system via deep reinforcement learning. In Proceedings of the 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), Yokohama, Japan, 16–19 October 2017; pp. 1–6.
49. Belletti, F.; Haziza, D.; Gomes, G.; Bayen, A.M. Expert Level control of Ramp Metering based on Multi-task Deep Reinforcement Learning. *arXiv* **2017**, arXiv:1701.08832.
50. Chalaki, B.; Beaver, L.E.; Remer, B.; Jang, K.; Vinitsky, E.; Bayen, A.M.; Malikopoulos, A.A. Zero-shot autonomous vehicle policy transfer: From simulation to real-world via adversarial learning. *arXiv* **2019**, arXiv:1903.05252.
51. 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.
52. Qi, X.; Luo, Y.; Wu, G.; Boriboonsomsin, K.; Barth, M. Deep reinforcement learning enabled self-learning control for energy efficient driving. *Transp. Res. Part Emerg. Technol.* **2019**, *99*, 67–81. [CrossRef]
53. 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.
54. 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. [CrossRef]
55. 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.
56. Holmgren, V. General-Purpose Maintenance Planning Using Deep Reinforcement Learning and Monte Carlo Tree Search. Master's Dissertation, Linköping University, Linköping, Sweden, December 2019.
57. 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.
58. Huang, J.; Chang, Q.; Arinez, J. Deep reinforcement learning based preventive maintenance policy for serial production lines. *Expert Syst. Appl.* **2020**, *160*, 113701. [CrossRef]
59. 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.
60. Andriotis, C.P.; Papakonstantinou, K.G. Managing engineering systems with large state and action spaces through deep reinforcement learning. *arXiv* **2018**, arXiv:1811.02052.
61. 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. [CrossRef]
62. 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.
63. 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. [CrossRef]
64. 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. [CrossRef]
65. 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. [CrossRef]

66. 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. [CrossRef]
67. 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. [CrossRef]
68. 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. [CrossRef]
69. 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. [CrossRef]
70. Yang, H.; Xie, X.; Kadoch, M. Machine Learning Techniques and A Case Study for Intelligent Wireless Networks. *IEEE Netw.* **2020**, *34*, 208–215. [CrossRef]
71. Lei, L.; Tan, Y.; Zheng, K.; Liu, S.; Zhang, K.; Shen, X. Deep Reinforcement Learning for Autonomous Internet of Things: Model, Applications and Challenges. *IEEE Commun. Surv. Tutor.* **2020**, *22*, 1722–1760. [CrossRef]
72. Liu, M.; Yu, F.R.; Teng, Y.; Leung, V.C.; Song, M. Performance optimization for blockchain-enabled industrial Internet of Things (IIoT) systems: A deep reinforcement learning approach. *IEEE Trans. Ind. Inform.* **2019**, *15*, 3559–3570. [CrossRef]
73. Liu, C.H.; Lin, Q.; Wen, S. Blockchain-Enabled Data Collection and Sharing for Industrial IoT With Deep Reinforcement Learning. *IEEE Trans. Ind. Inform.* **2019**, *15*, 3516–3526. [CrossRef]
74. He, Y.; Yu, F.R.; Zhao, N.; Leung, V.C.; Yin, H. Software-defined networks with mobile edge computing and caching for smart cities: A big data deep reinforcement learning approach. *IEEE Commun. Mag.* **2017**, *55*, 31–37. [CrossRef]
75. Bu, F.; Wang, X. A smart agriculture IoT system based on deep reinforcement learning. *Future Gener. Comput. Syst.* **2019**, *99*, 500–507. [CrossRef]
76. Zhao, R.; Wang, X.; Xia, J.; Fan, L. Deep reinforcement learning based mobile edge computing for intelligent Internet of Things. *Phys. Commun.* **2020**, *43*, 101184. [CrossRef]
77. Zhu, H.; Cao, Y.; Wei, X.; Wang, W.; Jiang, T.; Jin, S. Caching transient data for Internet of Things: A deep reinforcement learning approach. *IEEE Internet Things J.* **2018**, *6*, 2074–2083. [CrossRef]
78. Chen, J.; Chen, S.; Wang, Q.; Cao, B.; Feng, G.; Hu, J. iRAF: A deep reinforcement learning approach for collaborative mobile edge computing IoT networks. *IEEE Internet Things J.* **2019**, *6*, 7011–7024. [CrossRef]
79. Wei, Y.; Yu, F.R.; Song, M.; Han, Z. Joint optimization of caching, computing, and radio resources for fog-enabled IoT using natural actor–critic deep reinforcement learning. *IEEE Internet Things J.* **2018**, *6*, 2061–2073. [CrossRef]
80. Sun, Y.; Peng, M.; Mao, S. Deep reinforcement learning-based mode selection and resource management for green fog radio access networks. *IEEE Internet Things J.* **2018**, *6*, 1960–1971. [CrossRef]
81. Gazori, P.; Rahbari, D.; Nickray, M. Saving time and cost on the scheduling of fog-based IoT applications using deep reinforcement learning approach. *Future Gener. Comput. Syst.* **2020**, *110*, 1098–1115. [CrossRef]
82. Zhu, J.; Song, Y.; Jiang, D.; Song, H. A new deep-Q-learning-based transmission scheduling mechanism for the cognitive Internet of Things. *IEEE Internet Things J.* **2017**, *5*, 2375–2385. [CrossRef]
83. Ferdowsi, A.; Saad, W. Deep learning for signal authentication and security in massive internet-of-things systems. *IEEE Trans. Commun.* **2018**, *67*, 1371–1387. [CrossRef]
84. Jay, N.; Rotman, N.; Godfrey, B.; Schapira, M.; Tamar, A. A deep reinforcement learning perspective on internet congestion control. In Proceedings of the International Conference on Machine Learning, Long Beach, CA, USA, 10–15 June 2019; pp. 3050–3059.
85. He, Y.; Zhao, N.; Yin, H. Integrated networking, caching, and computing for connected vehicles: A deep reinforcement learning approach. *IEEE Trans. Veh. Technol.* **2017**, *67*, 44–55. [CrossRef]
86. Doddalinganavar, S.S.; Tergundi, P.V.; Patil, R.S. Survey on Deep Reinforcement Learning Protocol in VANET. In Proceedings of the 2019 1st International Conference on Advances in Information Technology (ICAIT), Chikmagalur, India, 25–27 July 2019; pp. 81–86. [CrossRef]
87. Tan, L.T.; Hu, R.Q. Mobility-aware edge caching and computing in vehicle networks: A deep reinforcement learning. *IEEE Trans. Veh. Technol.* **2018**, *67*, 10190–10203. [CrossRef]
88. Ning, Z.; Dong, P.; Wang, X.; Rodrigues, J.J.; Xia, F. Deep reinforcement learning for vehicular edge computing: An intelligent offloading system. *ACM Trans. Intell. Syst. Technol. (TIST)* **2019**, *10*, 1–24. [CrossRef]
89. Liu, Y.; Yu, H.; Xie, S.; Zhang, Y. Deep reinforcement learning for offloading and resource allocation in vehicle edge computing and networks. *IEEE Trans. Veh. Technol.* **2019**, *68*, 11158–11168. [CrossRef]
90. Ning, Z.; Zhang, K.; Wang, X.; Obaidat, M.S.; Guo, L.; Hu, X.; Hu, B.; Guo, Y.; Sadoun, B.; Kwok, R.Y.K. Joint Computing and Caching in 5G-Envisioned Internet of Vehicles: A Deep Reinforcement Learning-Based Traffic Control System. *IEEE Trans. Intell. Transp. Syst.* **2020**, 1–12. [CrossRef]
91. Ye, H.; Li, G.Y.; Juang, B.H.F. Deep reinforcement learning based resource allocation for V2V communications. *IEEE Trans. Veh. Technol.* **2019**, *68*, 3163–3173. [CrossRef]
92. Chinchali, S.; Hu, P.; Chu, T.; Sharma, M.; Bansal, M.; Misra, R.; Pavone, M.; Katti, S. Cellular Network Traffic Scheduling With Deep Reinforcement Learning. In Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018; pp. 766–774.
93. Kwon, D.; Kim, J. Multi-Agent Deep Reinforcement Learning for Cooperative Connected Vehicles. In Proceedings of the 2019 IEEE Global Communications Conference (GLOBECOM), Waikoloa, HI, USA, 9–13 December 2019; pp. 1–6.

94. Aoki, S.; Higuchi, T.; Altintas, O. Cooperative Perception with Deep Reinforcement Learning for Connected Vehicles. *arXiv* **2020**, arXiv:2004.10927.
95. Mao, H.; Alizadeh, M.; Menache, I.; Kandula, S. Resource management with deep reinforcement learning. In Proceedings of the 15th ACM Workshop on Hot Topics in Networks, Atlanta, GA USA, 9–10 November 2016; pp. 50–56.
96. Li, R.; Zhao, Z.; Sun, Q.; Chih-Lin, I.; Yang, C.; Chen, X.; Zhao, M.; Zhang, H. Deep reinforcement learning for resource management in network slicing. *IEEE Access* **2018**, *6*, 74429–74441. [CrossRef]
97. Zhang, Y.; Yao, J.; Guan, H. Intelligent cloud resource management with deep reinforcement learning. *IEEE Cloud Comput.* **2017**, *4*, 60–69. [CrossRef]
98. Esteva, A.; Robicquet, A.; Ramsundar, B.; Kuleshov, V.; DePristo, M.; Chou, K.; Cui, C.; Corrado, G.; Thrun, S.; Dean, J. A guide to deep learning in healthcare. *Nat. Med.* **2019**, *25*, 24–29. [CrossRef]
99. Liu, S.; Ngiam, K.Y.; Feng, M. Deep Reinforcement Learning for Clinical Decision Support: A Brief Survey. *arXiv* **2019**, arXiv:1907.09475.
100. Yu, C.; Liu, J.; Nemati, S. Reinforcement learning in healthcare: A survey. *arXiv* **2019**, arXiv:1908.08796.
101. Liu, Y.; Logan, B.; Liu, N.; Xu, Z.; Tang, J.; Wang, Y. Deep reinforcement learning for dynamic treatment regimes on medical registry data. In Proceedings of the 2017 IEEE International Conference on Healthcare Informatics (ICHI), Park City, UT, USA, 23–26 August 2017; pp. 380–385.
102. Reddy, S.; Levine, S.; Dragan, A. Accelerating human learning with deep reinforcement learning. In Proceedings of the NIPS 2017 Workshop: Teaching Machines, Robots, and Humans, Long Beach, CA, USA, 4–9 December 2017.
103. Zheng, G.; Zhang, F.; Zheng, Z.; Xiang, Y.; Yuan, N.J.; Xie, X.; Li, Z. DRN: A deep reinforcement learning framework for news recommendation. In Proceedings of the 2018 World Wide Web Conference, Lyon, France, 23–27 April 2018; pp. 167–176.
104. Ritter, G. Reinforcement learning in finance. In *Big Data and Machine Learning in Quantitative Investment*; John Wiley & Sons: Hoboken, NJ, USA, 2018; pp. 225–250.
105. Charpentier, A.; Elie, R.; Remlinger, C. Reinforcement Learning in Economics and Finance. *arXiv* **2020**, arXiv:2003.10014.
106. Xiong, Z.; Liu, X.Y.; Zhong, S.; Yang, H.; Walid, A. Practical deep reinforcement learning approach for stock trading. *arXiv* **2018**, arXiv:1811.07522.
107. Zhang, Z.; Zohren, S.; Roberts, S. Deep reinforcement learning for trading. *J. Financ. Data Sci.* **2020**, *2*, 25–40. [CrossRef]
108. Wang, J.; Zhang, Y.; Tang, K.; Wu, J.; Xiong, Z. Alphastock: A buying-winners-and-selling-losers investment strategy using interpretable deep reinforcement attention networks. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage, AK, USA, 4–8 August 2019; pp. 1900–1908.
109. Jiang, Z.; Xu, D.; Liang, J. A deep reinforcement learning framework for the financial portfolio management problem. *arXiv* **2017**, arXiv:1706.10059.
110. Hu, Y.J.; Lin, S.J. Deep reinforcement learning for optimizing finance portfolio management. In Proceedings of the 2019 Amity International Conference on Artificial Intelligence (AICAI), Dubai, United Arab Emirates, 4–6 February 2019; pp. 14–20.
111. Harris, A.T.; Schaub, H. Spacecraft Command and Control with Safety Guarantees using Shielded Deep Reinforcement Learning. In Proceedings of the AIAA Scitech 2020 Forum, Orlando, FL, USA, 6–10 January 2020; p. 0386.
112. Harris, A.; Teil, T.; Schaub, H. Spacecraft decision-making autonomy using deep reinforcement learning. In Proceedings of the 29th AAS/AIAA Space Flight Mechanics Meeting, Maui, HI, USA, 13–17 January 2019; pp. 1–19.
113. Yu, L.; Wang, Q.; Guo, Y.; Li, P. Spectrum availability prediction in cognitive aerospace communications: A deep learning perspective. In Proceedings of the 2017 Cognitive Communications for Aerospace Applications Workshop (CCAA), Cleveland, OH, USA, 27–28 June 2017; pp. 1–4.
114. Liu, C.H.; Chen, Z.; Tang, J.; Xu, J.; Piao, C. Energy-efficient UAV control for effective and fair communication coverage: A deep reinforcement learning approach. *IEEE J. Sel. Areas Commun.* **2018**, *36*, 2059–2070. [CrossRef]
115. Julian, K.D.; Kochenderfer, M.J. Distributed wildfire surveillance with autonomous aircraft using deep reinforcement learning. *J. Guid. Control Dyn.* **2019**, *42*, 1768–1778. [CrossRef]
116. Dulac-Arnold, G.; Mankowitz, D.; Hester, T. Challenges of real-world reinforcement learning. *arXiv* **2019**, arXiv:1904.12901.
117. 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.
118. Osband, I.; Blundell, C.; Pritzel, A.; Roy, B.V. Deep Exploration via Bootstrapped DQN. *arXiv* **2016**, arXiv:1602.04621.
119. Dulac-Arnold, G.; Evans, R.; Sunehag, P.; Coppin, B. Reinforcement Learning in Large Discrete Action Spaces. *arXiv* **2015**, arXiv:1512.07679.
120. 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.
121. 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.
122. Boutilier, C.; Lu, T. *Budget Allocation Using Weakly Coupled, Constrained Markov Decision Processes*; ResearchGate: Berlin, Germany, 2016.
123. 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.

124. 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.
125. Hester, T.; Quinlan, M.J.; Stone, P. A Real-Time Model-Based Reinforcement Learning Architecture for Robot Control. *arXiv* **2011**, arXiv:1105.1749.
126. 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. [CrossRef] [PubMed]
127. 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.
128. Arjona-Medina, J.A.; Gillhofer, M.; Widrich, M.; Unterthiner, T.; Hochreiter, S. RUDDER: Return Decomposition for Delayed Rewards. *arXiv* **2018**, arXiv:1806.07857.