entropy

MDPI

Editorial

Machine Learning for Communications

Vaneet Aggarwal

School of Industrial Engineering and School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907, USA; vaneet@purdue.edu

Due to the proliferation of applications and services that run over communication networks, ranging from video streaming and data analytics to robotics and augmented reality, tomorrow's networks will be faced with increasing challenges resulting from the explosive growth of data traffic demand with significantly varying performance requirements. This calls for more powerful, intelligent methods to enable novel network design, deployment, and management. To realize this vision, there is an increasing need to leverage recent developments in machine learning (ML), as well as other artificial intelligence (AI) techniques, and fully integrate them into the design and optimization of communication networks.

In this editorial, we will first summarize the key problem structures in communication systems where machine learning solutions have been used. Then, we will describe the areas where there are gaps in learning algorithms for their optimal applications to communication systems.

In the following, we will describe the different problem structures in communication systems, which can be solved by ML approaches.

Parametric Optimization with Deep Neural Networks: The formulation of parametric optimization is given as follows.

$$x^*(\theta) \quad = \quad \arg\min_x f(x,\theta) \tag{1}$$

In this problem, the aim is to represent the solution to an entire family of problems (for all θ). One approach for solving such problems is to use a Deep Neural Network with θ as an input and $x^*(\theta)$ as the output. Using a certain values of θ, the optimization problem can be solved and these values make the training data for the neural network. The trained neural network is then used to obtain $x^*(\theta)$ for all θ. Such approaches has been used for beamforming [1] as well as power control [2]. In these problems, the channel coefficients or the signal-to-noise ratio of the links are the parameters θ based on which optimal beamforming vectors or power control solution needs to be calculated.

We note that even for a given θ, finding $x^*(\theta)$ maybe a hard problem which limits obtaining enough training data for the problem. Recently, the authors of [2] proposed an approach where θ is sampled, and a single step along the gradient of the objective function is taken. This allows more flexibility as the optimization problem do not need to be fully solved for the training examples. Such an approach has been validated on power control problems by the authors of [2]. While such direction has great empirical evidence, convergence rates to global optimal $x^*(\theta)$ with samples is an open problem, to the best of our knowledge.

Reinforcement Learning for Combinatorial Optimization: Many problems in communication systems require combinatorial optimization, e.g., routing optimization, scheduling, and resource allocation [3]. Many combinatorial problems are NP-hard, and thus key approaches for such problems have been approximation algorithms, hand-crafted heuristics, or meta-heuristics [4]. The combinatorial optimization problem can be formulated as follows: Let $\mathcal{V}$ be a set of elements and $f : \mathcal{V} \to \mathbb{R}$ be a cost function. Combinatorial optimization problem aims to find an optimal value of the function f and any corresponding optimal element that achieves that optimal value on the domain. One of the emerging

Citation: Aggarwal, V. Machine Learning for Communications. *Entropy* **2021**, *23*, 831. https://doi.org/10.3390/e23070831

Received: 17 June 2021
Accepted: 21 June 2021
Published: 29 June 2021

recent trends is to solve combinatorial optimization problems by using a reinforcement learning approach. In this approach, the combinatorial optimization problem is formulated as a Markov Decision Process (MDP). The state encodes the current solution, and the action describes the modification of the current solution. The reward is given by the change in objective with the modification. The exact state and action encoding depends on the problem and the approach used. A recent survey of the different approaches based on reinforcement learning to combinatorial optimization are presented in [5]. We note that combinatorial optimization approaches using reinforcement learning have been used in communications to find efficient encoding designs [6–8].

Reinforcement Learning for Dynamic Resource Management: In the presence of dynamic job arrivals, online resource management of computing and communication resources become important. Consider an example of a single queue which is serving different types of customers. The overall objective is to minimize weighted latency of the different types of customers, where the queue needs to decide which of the customer request to be processed next. This can be modeled as a Markov decision process with the state as the vector composed of the queue length of each type of customer, action is to choose which of the customer request to be processed next, and the cost (negative of reward) is the weighted latency of the served customer. The current action impacts the next state and leads to a dynamic system. In networking problems, such scheduling problems occur at all layers, which make the use of reinforcement learning important in networking problems. In particular, modern networks such as Internet of Things (IoT), Heterogeneous Networks (HetNets), and Unmanned Aerial Vehicle (UAV) network become more decentralized, ad-hoc, and autonomous in nature. Network entities such as IoT devices, mobile users, and UAVs need to make local and autonomous decisions, e.g., spectrum access, data rate selection, transmit power control, and base station association, to achieve the goals of different networks including, e.g., throughput maximization, and energy consumption minimization [9]. This has led to widespread use of reinforcement learning in networking applications, see [9] for a detailed survey. Some of the applications include traffic engineering [10], caching [11], queue management [12], video streaming [13], software-defined networks [14]. In addition to wireless networks, reinforcement learning for dynamic resource management has been widely used in transportation networks, e.g., vehicle routing and dispatch [15–17], freight scheduling [18], and traffic signal control [19].

We will now describe some of the areas where novel learning-based solutions are needed, which have applications in communication research.

Joint Decision of Multiple Agents: Communication systems consist of multiple decision makers in the system, e.g., multiple base stations. With multiple decision makers, multiple challenges arise. One of them is that the joint decision requires joint state and joint action space of the users. However, this is computationally prohibitive. In order to deal with this challenge, multiple approaches have been proposed. One of the approaches is an approximation of cooperative multi-agent reinforcement leaning by a mean-field control (MFC) framework, where the approximation error is shown to be of $O(1/\sqrt{N})$ for N agents [20]. Another approach is the use of decentralizable algorithms, which aim to do centralized training and decentralized execution [21–23]. Further, there is a distributed approach which introduces communication among agents during execution [24,25]. Even though multiple approaches have been investigated, efficient complexity-performance-communication tradeoff is an important research problem.

Multi-objective Optimization: Many realistic applications have multiple objectives, e.g., capacity and power usage in the communication system [26,27], latency and energy consumption [28], efficiency and safety in robotic systems [29]. Further, the overall aim is to optimize a non-linear function of the different objectives. In this setup, standard reinforcement learning approaches do not work since the non-linear objective function looses the additive structure, and thus the Bellman's Equation does not work anymore in this setting [30]. Recently, this problem has been studied, where guarantees for model-based algorithm and model-free algorithm have been studied in [30,31], respectively. The ap-

proaches have been applied to cellular scheduling, traffic engineering, and queue scheduling problems. However, the research on this direction is still in its infancy, and scalable algorithms with better guarantees need investigation.

Constraints in Decision Making: Most communication systems have constraints, e.g., power, latency, etc. Consider a wireless sensor network where the devices aim to update a server with sensor values. At time t, the device can choose to send a packet to obtain a reward of 1 unit or to queue the packet and obtain 0 reward. However, communicating a packet results in p_t power consumption. At time t, if the wireless channel condition, s_t, is weak and the device chooses to send a packet, the resulting instantaneous power consumption, p_t, is high. The goal is to send as many packets as possible while keep the average power consumption, $\sum_{t=1}^{T} p_t/T$, within some limit, say C. This environment has state (s_t, q_t) as the channel condition and queue length at time t. To limit the power consumption, the agent may choose to send packets when the channel condition is good or when the queue length grows beyond a certain threshold. The agent aims to learn the policies in an *online manner* which requires efficiently balancing exploration of state-space and exploitation of the estimated system dynamics. Similar to the example above, many applications require to keep some costs low while simultaneously maximizing the rewards [32]. Some attempts to use constrained reinforcement learning approaches to communication problems can be seen in [12,33,34].

The problem setup, where the system dynamics are known, is extensively studied [32]. For a constrained setup, the optimal policy is possibly stochastic [32,35]. In the domain where the agent learns the system dynamics and aims to learn good policies online, there has been work where to show asymptotic convergence to optimal policies and regret guarantees for infinite horizon [36–38], as well as episodic MDPs [39,40]. Recently, guarantees for policy-gradient based approaches have been studied [41,42]. In addition, peak constraints have also been studied for convergence guarantees [43]. Further, algorithms with use of deep learning architectures have been studied [12,44]. Scalable algorithms with better guarantees in presence of constraints still need more investigation.

Adaptivity to changes in the environment: Most existing works on reinforcement learning consider a stationary environment and aim to find or be comparable to an optimal policy. In many applications, however, the environment is far from being stationary. As an example, network demands have diurnal patterns [45]. With dynamic changes in the environment, the strategies need to adapt. There has been two key approaches to measure non-stationarity of the environment. The first is where there are L changes in the system, and another is where the total amount of variation in the MDP is bounded by Δ. Different algorithms have been proposed to optimize the dynamic regret in this setup, with different amounts of information on L and Δ, for a comprehensive set of algorithms from regret perspective the reader is referred to [46]. Ideally, we require an adaptive algorithm that works without the knowledge of L and Δ, while achieving optimal regret bounds. Such algorithms have been shown in the episodic MDPs in tabular and linear cases [46]. There are partial results for infinite-horizon tabular case, while the proposed algorithm is not scalable. This is because the proposed algorithm opens multiple instances of base algorithms which increases the complexity of the approach. Recently, there has been an approach based on change point detection on the experience tuples to detect the change in MDPs [47], which has been applied to a sensor energy management problem and a traffic signal control problem in [47], and extended to adapt to diurnal patterns in demand of ride-sharing services in [48,49]. However, theoretical guarantees for such an approach are open.

Funding: This research received no external funding.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Xia, W.; Zheng, G.; Zhu, Y.; Zhang, J.; Wang, J.; Petropulu, A.P. A deep learning framework for optimization of MISO downlink beamforming. *IEEE Trans. Commun.* **2019**, *68*, 1866–1880. [CrossRef]
2. Nikbakht, R.; Jonsson, A.; Lozano, A. Unsupervised learning for parametric optimization. *IEEE Commun. Lett.* **2020**, *25*, 678–681. [CrossRef]
3. Cheng, M.X.; Li, Y.; Du, D.Z. *Combinatorial Optimization in Communication Networks*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2006; Volume 18.
4. Stefanello, F.; Aggarwal, V.; Buriol, L.S.; Resende, M.G. Hybrid algorithms for placement of virtual machines across geo-separated data centers. *J. Comb. Optim.* **2019**, *38*, 748–793. [CrossRef]
5. Mazyavkina, N.; Sviridov, S.; Ivanov, S.; Burnaev, E. Reinforcement learning for combinatorial optimization: A survey. *Comput. Oper. Res.* **2021**, *134*, 105400. [CrossRef]
6. Kim, H.; Jiang, Y.; Kannan, S.; Oh, S.; Viswanath, P. Deepcode: Feedback codes via deep learning. In Proceedings of the 32nd International Conference on Neural Information Processing Systems, Montreal, QC, Canada, 3–8 December 2018; pp. 9458–9468.
7. Huang, L.; Zhang, H.; Li, R.; Ge, Y.; Wang, J. AI coding: Learning to construct error correction codes. *IEEE Trans. Commun.* **2019**, *68*, 26–39. [CrossRef]
8. Chadaga, S.; Agarwal, M.; Aggarwal, V. Encoders and Decoders for Quantum Expander Codes Using Machine Learning. *arXiv* **2019**, arXiv:1909.02945.
9. Luong, N.C.; Hoang, D.T.; Gong, S.; Niyato, D.; Wang, P.; Liang, Y.C.; Kim, D.I. Applications of deep reinforcement learning in communications and networking: A survey. *IEEE Commun. Surv. Tutor.* **2019**, *21*, 3133–3174. [CrossRef]
10. Geng, N.; Lan, T.; Aggarwal, V.; Yang, Y.; Xu, M. A Multi-agent Reinforcement Learning Perspective on Distributed Traffic Engineering. In Proceedings of the 2020 IEEE 28th International Conference on Network Protocols (ICNP), Madrid, Spain, 13–16 October 2020; pp. 1–11.
11. Wang, Y.; Li, Y.; Lan, T.; Aggarwal, V. Deepchunk: Deep q-learning for chunk-based caching in wireless data processing networks. *IEEE Trans. Cogn. Commun. Netw.* **2019**, *5*, 1034–1045. [CrossRef]
12. Raghu, R.; Upadhyaya, P.; Panju, M.; Agarwal, V.; Sharma, V. Deep reinforcement learning based power control for wireless multicast systems. In Proceedings of the 2019 57th Annual Allerton Conference on Communication, Control, and Computing, Allerton, IL, USA, 24–27 September 2019; pp. 1168–1175.
13. Mao, H.; Netravali, R.; Alizadeh, M. Neural adaptive video streaming with pensieve. In Proceedings of the Conference of the ACM Special Interest Group on Data Communication, Los Angeles, CA, USA, 21–25 August 2017; pp. 197–210.
14. Zhang, J.; Ye, M.; Guo, Z.; Yen, C.Y.; Chao, H.J. CFR-RL: Traffic engineering with reinforcement learning in SDN. *IEEE J. Sel. Areas Commun.* **2020**, *38*, 2249–2259. [CrossRef]
15. Hildebrandt, F.D.; Thomas, B.; Ulmer, M.W. Where the Action is: Let's make Reinforcement Learning for Stochastic Dynamic Vehicle Routing Problems work! *arXiv* **2021**, arXiv:2103.00507.
16. Al-Abbasi, A.O.; Ghosh, A.; Aggarwal, V. Deeppool: Distributed model-free algorithm for ride-sharing using deep reinforcement learning. *IEEE Trans. Intell. Transp. Syst.* **2019**, *20*, 4714–4727. [CrossRef]
17. Haliem, M.; Mani, G.; Aggarwal, V.; Bhargava, B. A distributed model-free ride-sharing approach for joint matching, pricing, and dispatching using deep reinforcement learning. *arXiv* **2020**, arXiv:2010.01755.
18. Chen, J.; Umrawal, A.K.; Lan, T.; Aggarwal, V. DeepFreight: A Model-free Deep-reinforcement-learning-based Algorithm for Multi-transfer Freight Delivery. In Proceedings of the International Conference on Automated Planning and Scheduling, Guangzhou, China, 7–12 June 2021; Volume 31, pp. 510–518.
19. Chen, C.; Wei, H.; Xu, N.; Zheng, G.; Yang, M.; Xiong, Y.; Xu, K.; Li, Z. Toward A thousand lights: Decentralized deep reinforcement learning for large-scale traffic signal control. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 3414–3421.
20. Gu, H.; Guo, X.; Wei, X.; Xu, R. Mean-Field Controls with Q-learning for Cooperative MARL: Convergence and Complexity Analysis. *arXiv* **2020**, arXiv:2002.04131.
21. Rashid, T.; Samvelyan, M.; Schroeder, C.; Farquhar, G.; Foerster, J.; Whiteson, S. Qmix: Monotonic value function factorisation for deep multi-agent reinforcement learning. In Proceedings of the International Conference on Machine Learning (PMLR), Stockholm, Sweden, 10–15 June 2018; pp. 4295–4304.
22. Rashid, T.; Farquhar, G.; Peng, B.; Whiteson, S. Weighted QMIX: Expanding Monotonic Value Function Factorisation. *arXiv* **2020**, arXiv:2006.10800.
23. Zhang, J.; Bedi, A.S.; Wang, M.; Koppel, A. MARL with General Utilities via Decentralized Shadow Reward Actor-Critic. *arXiv* **2021**, arXiv:2106.00543.
24. Sukhbaatar, S.; Szlam, A.; Fergus, R. Learning multiagent communication with backpropagation. In Proceedings of the 30th International Conference on Neural Information Processing Systems, Barcelona, Spain, 5–10 December 2016; pp. 2252–2260.
25. Foerster, J.N.; Assael, Y.M.; de Freitas, N.; Whiteson, S. Learning to communicate with Deep multi-agent reinforcement learning. In Proceedings of the 30th International Conference on Neural Information Processing Systems, Barcelona, Spain, 5–10 December 2016; pp. 2145–2153.
26. Wang, Z.; Aggarwal, V.; Wang, X. Iterative dynamic water-filling for fading multiple-access channels with energy harvesting. *IEEE J. Sel. Areas Commun.* **2015**, *33*, 382–395. [CrossRef]

27. Aggarwal, V.; Bell, M.R.; Elgabli, A.; Wang, X.; Zhong, S. Joint energy-bandwidth allocation for multiuser channels with cooperating hybrid energy nodes. *IEEE Trans. Veh. Technol.* **2017**, *66*, 9880–9889. [CrossRef]
28. Badita, A.; Parag, P.; Aggarwal, V. Optimal Server Selection for Straggler Mitigation. *IEEE/ACM Trans. Netw.* **2020**, *28*, 709–721. [CrossRef]
29. Nishimura, M.; Yonetani, R. L2B: Learning to Balance the Safety-Efficiency Trade-off in Interactive Crowd-aware Robot Navigation. *arXiv* **2020**, arXiv:2003.09207.
30. Agarwal, M.; Aggarwal, V. Reinforcement Learning for Joint Optimization of Multiple Rewards. *arXiv* **2021**, arXiv:1909.02940v3.
31. Bai, Q.; Agarwal, M.; Aggarwal, V. Joint Optimization of Multi-Objective Reinforcement Learning with Policy Gradient Based Algorithm. *arXiv* **2021**, arXiv:2105.14125.
32. Altman, E. *Constrained Markov Decision Processes*; CRC Press: Boca Raton, FL, USA, 1999; Volume 7.
33. Li, H.; Wan, Z.; He, H. Constrained EV charging scheduling based on safe deep reinforcement learning. *IEEE Trans. Smart Grid* **2019**, *11*, 2427–2439. [CrossRef]
34. Zhang, Y.; Vuong, Q.; Ross, K.W. First order optimization in policy space for constrained deep reinforcement learning. *arXiv* **2020**, arXiv:2002.06506.
35. Puterman, M.L. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*; John Wiley & Sons: Hoboken, NJ, USA, 2014.
36. Gattami, A.; Bai, Q.; Aggarwal, V. Reinforcement Learning for Constrained Markov Decision Processes. In Proceedings of the International Conference on Artificial Intelligence and Statistics (PMLR), Virtual Conference, 13–15 April 2021; pp. 2656–2664.
37. Singh, R.; Gupta, A.; Shroff, N.B. Learning in Markov decision processes under constraints. *arXiv* **2020**, arXiv:2002.12435.
38. Agarwal, M.; Bai, Q.; Aggarwal, V. Markov Decision Processes with Long-Term Average Constraints. *arXiv* **2021**, arXiv:2106.06680.
39. Zheng, L.; Ratliff, L. Constrained upper confidence reinforcement learning. In Proceedings of the 2nd Conference on Learning for Dynamics and Control (PMLR), Berkeley, CA, USA, 11–12 June 2020; pp. 620–629.
40. Ding, D.; Wei, X.; Yang, Z.; Wang, Z.; Jovanovic, M. Provably efficient safe exploration via primal-dual policy optimization. In Proceedings of the International Conference on Artificial Intelligence and Statistics (PMLR), Virtual Conference, 13–15 April 2021; pp. 3304–3312.
41. Xu, T.; Liang, Y.; Lan, G. A Primal Approach to Constrained Policy Optimization: Global Optimality and Finite-Time Analysis. *arXiv* **2020**, arXiv:2011.05869.
42. Ding, D.; Zhang, K.; Basar, T.; Jovanovic, M. Natural Policy Gradient Primal-Dual Method for Constrained Markov Decision Processes. In Proceedings of the 34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, BC, Canada, 6–12 December, 2020.
43. Bai, Q.; Aggarwal, V.; Gattami, A. Provably Efficient Model-Free Algorithm for MDPs with Peak Constraints. *arXiv* **2020**, arXiv:2003.05555.
44. Liu, C.; Geng, N.; Aggarwal, V.; Lan, T.; Yang Y.; Xu, M. CMIX: Deep Multi-agent Reinforcement Learning with Peak and Average Constraints. In Proceedings of the 2021 European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD 2021), Virtual Conference, 13–17 September 2021.
45. Aggarwal, V.; Mahimkar, A.; Ma, H.; Zhang, Z.; Aeron, S.; Willinger, W. Inferring smartphone service quality using tensor methods. In Proceedings of the 2016 12th International Conference on Network and Service Management (CNSM), Montreal, QC, Canada, 31 October–4 November 2016; pp. 263–267.
46. Wei, C.Y.; Luo, H. Non-stationary Reinforcement Learning without Prior Knowledge: An Optimal Black-box Approach. *arXiv* **2021**, arXiv:2102.05406.
47. Padakandla, S.; Prabuchandran, K.J.; Bhatnagar, S. Reinforcement learning algorithm for non-stationary environments. *Appl. Intell.* **2020**, *50*, 3590–3606. [CrossRef]
48. Haliem, M.; Aggarwal, V.; Bhargava, B. AdaPool: An Adaptive Model-Free Ride-Sharing Approach for Dispatching using Deep Reinforcement Learning. In Proceedings of the 7th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation, Virtual Conference, 18–20 November 2020; pp. 304–305.
49. Haliem, M.; Aggarwal, V.; Bhargava, B. AdaPool: A Diurnal-Adaptive Fleet Management Framework using Model-Free Deep Reinforcement Learning and Change Point Detection. *arXiv* **2021**, arXiv:2104.00203.