[1]Dr Prasad Rayi, Akkisetti

Vn Hanuman, Gedela Naveena, Mamatha B

Reinforcement Learning for Optimizing Wireless Networks

JES Journal of Electrical Systems

Abstract:

The latest advancements in reinforcement learning (RL) have enhanced the implementation of online RL for wireless radio resource management (RRM). Nonetheless, online reinforcement learning algorithms need direct engagement with the environment, which may be unfavorable owing to the possible decline in performance resulting from the inevitable exploration inherent in reinforcement learning. This study first examines the use of offline reinforcement learning methods to address the RRM challenge. We assess several cutting-edge offline reinforcement learning algorithms, including behavior constrained Q-learning (BCQ), conservative Q-learning (CQL), and implicit Q-learning (IQL), for a particular radio resource management issue that seeks to optimize a linear combination of sum and 5-percentile rates via user scheduling. The efficacy of offline reinforcement learning for the resource allocation management issue is significantly influenced by the behavior policy used during data collection. We further provide an innovative offline reinforcement learning approach that utilizes heterogeneous datasets gathered from various behavior rules. We demonstrate that an appropriate combination of datasets allows offline reinforcement learning to provide a nearly optimum reinforcement learning policy, despite the significant suboptimality of all participating behavior policies.

Index Terms—Radio Resource Management, Offline Reinforcement Learning, Deep Reinforcement Learning.

Introduction

There is an increasing interest in using reinforcement learning (RL) to address radio resource management (RRM) challenges in wireless networks. The distinctive characteristics of wireless RRM are propelling this emerging trend. Initially, many RRM activities are sequential, including a resource allocation decision, subsequent observation of network performance, and feedback to the decision maker for policy adjustment. Secondly, real-world wireless network optimization challenges are sometimes too intricate to be represented as basic optimization problems, necessitating model-free solutions that can adapt to the unknown deployment.

Third, contemporary wireless networks provide well-defined control and feedback systems, facilitating the observation of system states and the collection of performance metrics.

These characteristics have initiated considerable endeavors in the advancement of reinforcement learning systems for wireless radio resource management. Section II provides an overview of related studies. The majority, if not all, of the current studies use online reinforcement learning, whereby the reinforcement learning policy progressively enhances via interaction with the environment without any pre-deployment data. The investigation of the previously uncharted environment, The endeavors of KY and CS receive some funding from the US National Science Foundation under grants CNS-2002902, ECCS-2029978, and SII-2132700. The research conducted by JY is partially funded by the US NSF under grants CNS-2003131, ECCS-2030026, and ECCS-

[1] [1,2,3,4]International School Of Technology And Sciences For Women, A.P, India.

2143559. This is the camera-ready version of the paper accepted at Asilomar 2023, highlighting that during the initial phases, when environmental information is limited and reinforcement learning exploration is predominantly stochastic, it is a crucial element for online reinforcement learning. However, it also represents a significant barrier to the implementation of cutting-edge reinforcement learning algorithms in practical wireless networks. The absence of performance assurance during reinforcement learning exploration implies that network users may momentarily experience suboptimal Quality of Service (QoS) to enable the learning agent to acquire knowledge on the deployment for a possibly improved reinforcement learning strategy. This tradeoff is, however, unfavorable for the wireless network operator as compared to model-based or rule-based solutions, which, while they may not attain the same level of performance as online reinforcement learning post-convergence, do not experience potentially substantial initial performance decline. This study advocates the use of offline reinforcement learning [1] for the optimization of wireless networks. Offline reinforcement learning seeks to train reinforcement learning agents with pre-existing datasets, therefore entirely circumventing the need for online interactions. This paradigm is especially appropriate for wireless Radio Resource Management, since wireless operators have implemented policies governing resource allocation and possess established means for gathering operational data. We examine the viability and efficacy of offline reinforcement learning (RL) for wireless resource management (RRM) by assessing advanced offline RL algorithms, including behavior constrained Q-learning (BCQ) [2], conservative Q-learning (CQL) [3], and implicit Q-learning (IQL) [4], in the context of a wireless user scheduling challenge aimed at optimizing a linear amalgamation of sum and 5-percentile rates. The efficacy of offline reinforcement learning (RL) is illustrated through comprehensive system simulations, revealing that the performance of offline RL in the user scheduling problem is heavily contingent upon the behavior policy employed for data collection; datasets derived from suboptimal behavior policies do not yield effective RL policies. To address this issue, we offer an innovative offline reinforcement learning system that utilizes heterogeneous datasets gathered via several behavioral regulations. Surprisingly, the amalgamation of datasets gathered from many behavior policies enables offline reinforcement learning to provide a near-optimal policy, despite the fact that all participating behavior rules are significantly inferior. We also examine the likely causes for the advantages of mixed datasets in offline reinforcement learning and provide relevant avenues for further study.

Background On Wireless Networks

Wireless networks have become essential to the contemporary digital environment, supporting many applications that facilitate daily activities and technical progress. These networks provide communication without physical links, use radio frequencies to convey data across varying distances, from short-range connections in local area networks (LANs) to vast connections in wide area networks (WANs). The widespread use of wireless technologies, such as cellular networks, Wi-Fi, and satellite communication, underscores their essential function in facilitating mobility, improving connection, and underpinning the Internet of Things (IoT). The proliferation of linked devices, fueled by the growth of smart technology and high-bandwidth applications, has intensified the difficulty of maintaining and optimizing wireless networks. Network operators have the formidable challenge of ensuring that these systems can accommodate increasing data traffic, maintain reliable connection, and adjust to swiftly changing circumstances.

Literature Survey

The literature on wireless network optimization indicates a transition from conventional static techniques to more dynamic, data-driven strategies. Conventional methods, dependent on static algorithms and rule-based approaches, often fail to accommodate the dynamic and intricate characteristics of contemporary wireless networks. Machine Learning (ML) methodologies have arisen as a notable innovation, providing improved adaptability and efficiency. Machine learning algorithms, including supervised, unsupervised, and reinforcement learning techniques, provide robust instruments for forecasting traffic patterns, identifying abnormalities, and optimizing resource distribution in real-time. Comparative analyses indicate that machine learning-based technologies often surpass conventional techniques in performance and flexibility. Recent breakthroughs underscore the amalgamation of machine learning with technologies such as network function virtualization and edge computing, ensuring enhanced responsiveness and efficiency in network administration.

A. Conventional Optimization Methods in Wireless Networks:

Conventional optimization methods in wireless networks have mostly depended on static algorithms and rule-based strategies. These approaches often emphasize predetermined strategies for the management of network resources, such as static routing. Vikram Nattamai Sankaran et al. / ESP JETA 3(3), 64-77, 2023 Protocols and fixed bandwidth allotment. Classical routing methods such as Dijkstra's or Bellman-Ford are intended to identify the shortest route in a network, supposing static circumstances that seldom represent the dynamic characteristics of real-world wireless networks. Moreover, techniques like Quality of Service (QoS) management and traffic engineering seek to optimize resource use by prioritizing traffic and regulating bandwidth according to established protocols. Although these methodologies have been fundamental in network management, their efficacy is limited by their lack of adaptability to swiftly changing network circumstances, resulting in inferior performance during heavy traffic times or network abnormalities. The fixed nature of these approaches often leads to inefficiencies, especially in settings marked by significant mobility and fluctuating interference levels.

B. Machine Learning Approaches in Wireless Network Optimization:

In recent years, Machine Learning (ML) has become a revolutionary tool in wireless network optimization. Machine learning methodologies provide a dynamic alternative to conventional techniques by using data-driven models that may learn from network circumstances and change in real time. A substantial amount of research has investigated the use of supervised learning techniques, including neural networks and support vector machines, to forecast network traffic patterns and enhance resource allocation. These models use previous data to predict future network needs and modify settings appropriately, so enhancing overall efficiency. Unsupervised learning methods, including as clustering and anomaly detection, have been used to discover and address atypical network activity, like interference or failures, without requiring labeled data. Moreover, reinforcement learning has been used to create adaptive methods that perpetually learn and enhance their behaviors depending on input from the network environment. The capacity for real-time modifications based on observed performance differentiates machine learning approaches from conventional optimization strategies, providing significant advancements in network resource management and connection enhancement.

C. Comparative Studies on ML and Traditional Methods:

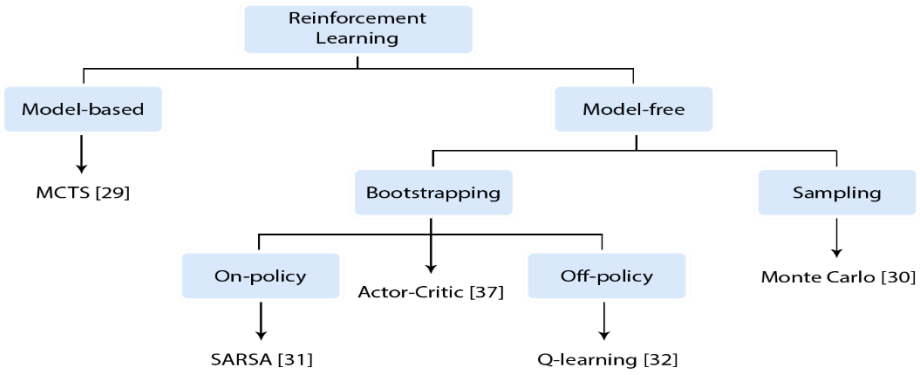
Comparative analyses of ML-based and classical optimization techniques have shown the benefits of integrating ML into network management. Studies comparing these methodologies have repeatedly shown that machine learning algorithms can surpass conventional techniques in efficiency and flexibility. Research has shown that machine learning models, particularly those using deep learning methodologies, may more effectively manage the fluctuating characteristics of network traffic by perpetually assimilating new data and refining their predictions. Conversely, conventional approaches often depend on static models that may fail to accommodate abrupt changes in network circumstances. Moreover, machine learning-based techniques have shown enhanced anomaly detection accuracy, minimizing false positives and augmenting the network's responsiveness to concerns promptly. Nevertheless, it is acknowledged that incorporating machine learning into network management has problems, including the need for substantial training data and processing resources. Notwithstanding these limitations, the advantages of increased flexibility and superior performance measures make machine learning a persuasive choice for contemporary wireless network optimization.

D. Recent Advances and Emerging Trends:

Recent advancements in machine learning and its application to wireless network optimization indicate an increasing interest in using advanced technology to tackle intricate network management challenges. Current developments include the amalgamation of machine learning with network function virtualization (NFV) and software-defined networking (SDN), enabling more adaptable and programmable network designs. These technologies provide the dynamic allocation of resources and the deployment of machine learning algorithms across several network levels, enhancing network management efficiency and responsiveness. Moreover, there is a growing emphasis on real-time data processing and edge computing, which seek to position machine learning capabilities nearer to data sources, therefore minimizing latency and enhancing decision-making speed. Research is investigating the use of sophisticated machine learning approaches, including transfer learning and meta-learning, to enhance model efficacy in scenarios characterized by little data or dynamic situations. These developments highlight the capacity of machine learning to foster innovation in wireless network optimization and tackle growing difficulties in network management.

Overview of Reinforcement Learning

This section offers an extensive overview of several reinforcement learning approaches, including the principles of Markov Decision Processes, Partially Observable Markov Decision Processes (POMDP), and Deep Reinforcement Learning models along with their distinct characteristics.



Machine Learning Techniques

Machine Learning (ML) provides a variety of methodologies that may be customized to enhance several facets of wireless network performance, such as data flow management and connection. This section examines several fundamental machine learning techniques—supervised learning, unsupervised learning, reinforcement learning, and hybrid approaches—emphasizing their applications, advantages, and roles in improving wireless networks.

A. Supervised Learning Techniques:

Supervised learning approaches include training models on labeled datasets to forecast outcomes or categorize data based on established inputs. These strategies are especially beneficial in situations when historical data is accessible and can be used to train models that predict future network conditions or behaviors.

Reinforcement Learning Techniques

Reinforcement Learning (RL) is instructing an agent to make choices via interaction with an environment, receiving rewards or penalties contingent upon its actions. Reinforcement Learning is very adept at managing dynamic and adaptable network activities.

a) Q-Learning and Deep Q-Networks (DQNs):

Q-Learning is a model-free reinforcement learning method that determines the value of actions based on input from the environment. It is used to formulate rules for resource distribution and network optimization by maximizing aggregate rewards. Deep Q-Networks integrate Q-Learning with deep learning, enabling the management of high-dimensional state spaces and intricate decision-making situations. For example, DQNs may enhance bandwidth allocation and routing techniques by analyzing real-time network performance data.

b) Policy Gradient Methods and Actor-Critic Algorithms:

Policy Gradient approaches directly optimize the policy by modifying the parameters that govern the agent's activities. These strategies are advantageous for intricate decision-making challenges when the action space is either continuous or extensive. Actor-Critic algorithms integrate policy gradient techniques with value function estimation, facilitating effective learning and adaptability. These methodologies may be used for dynamic network management activities, including adaptive traffic control and QoS optimization, necessitating ongoing modifications in response to changing network circumstances.

D. Hybrid and Ensemble Approaches:

Hybrid and ensemble methods integrate many machine learning techniques to exploit their complimentary advantages, leading to enhanced performance and resilience.

a) Hybrid Models:

Hybrid models amalgamate diverse machine learning approaches to tackle distinct facets of network optimization. A hybrid model may use supervised learning to forecast traffic patterns and utilize unsupervised learning for anomaly identification within those patterns. This integration facilitates a holistic approach to network management by merging predictive capabilities with anomaly detection to improve overall network performance.

The dataset demonstrates that offline reinforcement learning enables the system to use the benefits of reinforcement learning without direct engagement with the environment. This is facilitated by the use of offline datasets. The predominant method for acquiring such datasets is gathering operational data linked to the current policy. For instance, in the context of the wireless RRM issue, wireless operators often possess pre-existing solutions that have been implemented in the intended environment.

The data gathered by the current solutions, referred to as behavior policies (BPs), may be used to train an offline reinforcement learning policy.

In the user scheduling issue, we have implemented four rule-based policies outlined in Section IV-B as BPs. Furthermore, we include two additional BPs that are founded on online RL.

The two strategies vary in their training efficacy; one is terminated at epoch 125, while the other concludes at epoch 350. Their achievements are delineated by the yellow dashed line and the black dashed line in Fig. 3. Utilizing datasets obtained from these BPs, we delineate the offline RL experimental technique as follows. Select a BP $\pi\beta$ from the available BPs.

Execute $\pi\beta$ in the environment to gather a dataset $D\pi\beta$.

3) Employ an offline reinforcement learning method to train policy $\pi\theta$ using the dataset $D\pi\beta$.

We note that the dataset gathered in Step 2 may exhibit subpar quality due to the associated BP potentially underperforming. For instance, as seen in Fig. 3, the four rule-based policies exhibit considerable performance disparities relative to the proficiently taught online reinforcement learning. We want to assess the feasibility of training an effective reinforcement learning policy using datasets derived from suboptimal behavioral policies.

Conclusion

We have presented offline reinforcement learning as a viable solution for the wireless radio resource management issue. To achieve this objective, we used a particular wireless user scheduling issue as a case study and assessed several cutting-edge offline reinforcement learning algorithms regarding their long-term efficacy and convergence rate. We noted that the efficacy of offline reinforcement learning is significantly limited by the quality of the behavior policy used to gather the dataset, and subsequently developed an innovative offline reinforcement learning method that utilizes heterogeneous datasets obtained from diverse behavior rules. We demonstrated that this technique is effective. Note that online RL-125 and ITLinQ exhibit comparable performance; hence, their allocations are same may generate a nearly optimum reinforcement learning policy in the presence of significantly poor behavior regulations, and offered many potential causes. Our research used a centralized offline reinforcement learning system to generate the RRM policy, which exhibits poor scalability. Developing a multi-agent offline reinforcement learning solution for this issue would be intriguing.

References

[1] S. Levine, A. Kumar, G. Tucker, and J. Fu, “Offline reinforcement learning: Tutorial, review, and perspectives on open problems,” arXiv preprint arXiv:2005.01643, 2020.

[2] S. Fujimoto, D. Meger, and D. Precup, “Off-policy deep reinforcement learning without exploration,” in International Conference on Machine Learning. PMLR, 2019, pp. 2052–2062.

[3] A. Kumar, A. Zhou, G. Tucker, and S. Levine, “Conservative Q-learning for offline reinforcement learning,” Advances in Neural Information Processing Systems, vol. 33, pp. 1179–1191, 2020.

[4] I. Kostrikov, A. Nair, and S. Levine, "Offline reinforcement learning with implicit Q-learning," arXiv preprint arXiv:2110.06169, 2021.

[5] K. I. Ahmed and E. Hossain, "A deep Q-learning method for downlink power allocation in multi-cell networks," arXiv preprint arXiv:1904.13032, 2019.

[6] F. Meng, P. Chen, and L. Wu, "Power allocation in multi-user cellular networks with deep Q learning approach," in IEEE International Conference on Communications (ICC). IEEE, 2019, pp. 1–6.

[7] G. Zhao, Y. Li, C. Xu, Z. Han, Y. Xing, and S. Yu, "Joint power control and channel allocation for interference mitigation based on reinforcement learning," IEEE Access, vol. 7, pp. 177 254–177 265, 2019.

[8] K. Yang, C. Shen, and T. Liu, "Deep reinforcement learning based wireless network optimization: A comparative study," in IEEE INFOCOM Workshop on Data Driven Intelligence for Networks, Toronto, Canada,Jul. 2020, pp. 1248–1253.

[9] Y. S. Nasir and D. Guo, "Multi-agent deep reinforcement learning for dynamic power allocation in wireless networks," IEEE Journal on Selected Areas in Communications, vol. 37, no. 10, pp. 2239–2250, 2019.

[10] N. Naderializadeh, J. J. Sydir, M. Simsek, and H. Nikopour, "Resource management in wireless networks via multi-agent deep reinforcement learning," IEEE Transactions on Wireless Communications, vol. 20, no. 6, pp. 3507–3523, 2021.

[11] P. Rashidinejad, B. Zhu, C. Ma, J. Jiao, and S. Russell, "Bridging offline reinforcement learning and imitation learning: A tale of pessimism," Advances in Neural Information Processing Systems, vol. 34, pp. 11 702–11 716, 2021.

[12] T. Xie, N. Jiang, H. Wang, C. Xiong, and Y. Bai, "Policy finetuning: Bridging sample-efficient offline and online reinforcement learning," Advances in Neural Information Processing Systems, vol. 34, pp. 27 395–27 407, 2021.

[13] A. Kumar, J. Hong, A. Singh, and S. Levine, "When should we prefer offline reinforcement learning over behavioral cloning?" arXiv preprint arXiv:2204.05618, 2022.

[14] 3GPP, "Simulation assumptions and parameters for FDD HeNB RF requirements," Tech. Rep. R4-092042.

[15] N. Mastronarde and M. van der Schaar, "Joint physical-layer and systemlevel power management for delay-sensitive wireless communications," IEEE Transactions on Mobile Computing, vol. 12, no. 4, pp. 694–709,2012.

[16] Z. Lin and M. van der Schaar, "Autonomic and distributed joint routing and power control for delay-sensitive applications in multi-hop wireless networks," IEEE Transactions on Wireless Communications, vol. 10, no. 1, pp. 102–113, 2010