DROO: Integrated Learning and Optimization for Edge Computing Offloading

Qian Zhang, Hong Kong University of Science and Technology

This installment of Computer's *series highlighting the work published in IEEE Computer Society journals comes from* IEEE Transactions on Mobile Computing.

Edge computing is recognized as a key enabling technology for Internet of Things (IoT) applications to achieve energy-efficient, ultralow-latency, and high-performance computing service beyond 5G/6G. IoT devices of limited computing power and onboard battery life can significantly improve their data processing capability by offloading intensive computation tasks to a nearby edge server for remote execution. Constrained by the available computation and communication resources in an edge access network, efficient provisioning of edge computing services requires judiciously deciding the set of IoT devices allowed to offload as well as the assets allocated to them based on real-time edge parameters. Achieving optimal online provisioning of edge services is challenging in a large IoT network under fast-varying wireless channel conditions. In particular, it requires repeatedly solving mixed-integer nonlinear programming (MINLP) problems, each to be finished within a channel coherence time of at most several seconds. Conventional model-based integer optimization methods are too slow in their numerical iterations to search large solution vector spaces, while model-free learning-based methods suffer from slow learning convergence or even divergence in finding an optimal hybrid integer-continuous online policy.

In "Deep Reinforcement Learning for Online Computation Offloading in Wireless Powered Mobile–Edge

Digital Object Identifier 10.1109/MC.2021.3055923
Date of current version: 17 November 2021

0018-9162/21©2021IEEE

Computing Networks," Huang et al.[1] propose a novel integrated optimization and learning approach, deep reinforcement learning-based online offloading (DROO). That can provision close-to-optimal computing services that are adaptive to fast-varying wireless channels. As shown in Figure 1, instead of solving for the hybrid integer-continuous solution altogether, DROO decomposes the original optimization problem into a zero–one binary offloading decision subproblem and a continuous resource allocation subproblem and then tackles them separately through a model-free learning module and a model-based optimization module, respectively. In each time slot, the actor module uses a fully connected deep neural network (DNN) to map the input instant system parameters to multiple binary offloading action vectors, and the critic module optimizes the resource allocation for each action vector and selects the best one. DROO uses the selected binary offloading action and the corresponding resource allocation as the control decision in the current time slot and then stores the selected binary action in a replay buffer for updating the policy of the DNN in the future. With a new input channel condition in the next time slot, the update of the DNN model parameters repeats until converging to the optimal online policy.

> To better approximate the submanifold, their method considers only a few basis modes in the vicinity of the desired deformation.

Compared with model-based optimization, DROO takes negligible computation time, as it can directly map any new channel input to the optimal output control action without the need for numerical optimization. On the other hand, compared with conventional deep reinforcement learning (DRL) methods that treat both the integer offloading and continuous resource allocation decisions as the action, DROO significantly reduces the action space of the DNN, which greatly simplifies the learning task to a classical classification problem. Perhaps more importantly, the optimization-based critic module provides precise evaluation of the integer offloading decisions generated by the actor module. This greatly improves the convergence of the training process as compared with conventional DRL, whose convergence is often jeopardized by the inaccurate evaluation of actions before the critic network is sufficiently trained. Simulation results show that DROO quickly converges in fewer than 3,000 iterations in a 30-user network, and the provisioned computing service achieves 99.9% of the optimum in less than 0.1 s after convergence.

Besides edge service provisioning, DROO has important applications in a wide range of areas, such as wireless communications and industrial control, where MINLP instances need to be frequently resolved. In addition, its decoupled structure facilitates simple modifications to handle more complicated

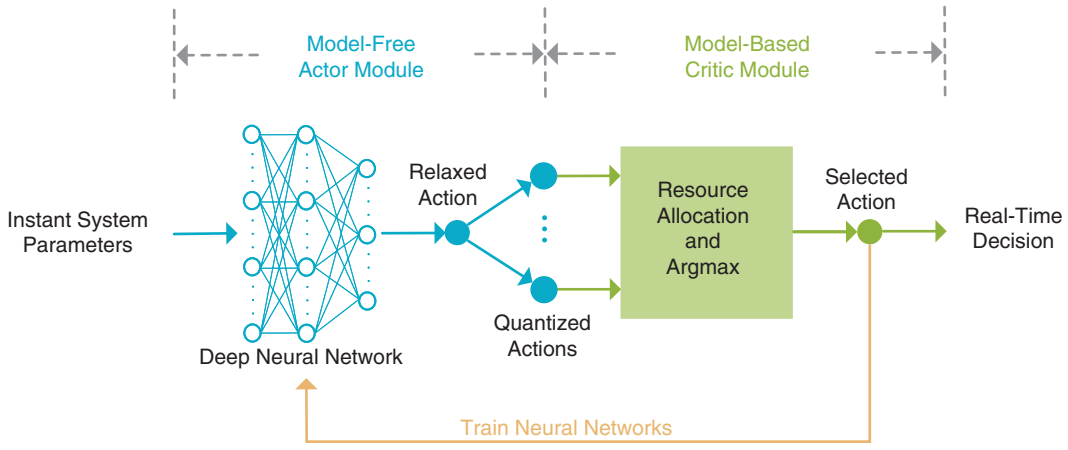

FIGURE 1. A schematic of DROO.[1]

use cases. For instance, by slightly modifying the critic module with DROO makes it truly viable to obtain real-time and optimal solutions to a variety of hard MINLP problems in fast-varying environments.

> Model-free learning-based methods suffer from slow learning convergence or even divergence in finding an optimal hybrid integer-continuous online policy.

a Lyapunov optimization technique, DROO can effectively tackle stochastic MINLP with long-term performance constraints[2]; by replacing the DNN in the actor module with a convolutional neural network, recursive neural network, or state-of-the-art transformer, DROO can improve its convergence performance and handle more sophisticated learning tasks. In this sense,

QIAN ZHANG is the Tencent Professor of Engineering and chair professor in the Department of Computer Science and Engineering, Hong Kong University of Science and Technology, Hong Kong, China. Contact him at qianzh@cse.ust.hk.

REFERENCES

1. L. Huang, S. Bi, and Y. J. Zhang, "Deep reinforcement learning for online computation offloading in wireless powered mobile-edge computing networks," *IEEE Trans. Mobile Comput.*, vol. 19, no. 11, pp. 2581–2593, Nov. 2020. doi: 10.1109/TMC.2019.2928811.
2. S. Bi, L. Huang, H. Wang, and Y. J. Zhang, "Lyapunov-guided deep reinforcement learning for stable online computation offloading in mobile-edge computing networks," *IEEE Trans. Wireless Commun.*, 2021, arXiv:2010.01370.

Digital Object Identifier 10.1109/MC.2021.3122778