JDS

Batched Neural Bandits

QUANQUAN GU, University of California, Los Angeles, USA
AMIN KARBASI, Yale University, USA
KHASHAYAR KHOSRAVI, Google Research NYC, USA
VAHAB MIRROKNI, Google Research NYC, USA
DONGRUO ZHOU, University of California, Los Angeles, USA

PROBLEM STATEMENT

We study the stochastic contextual bandit problem under the batched setting, where the agent is only allowed to change her policy at the end of each batch. Existing batched bandit methods focus on problems with simple reward structures, such as multi-armed bandits or linear bandits. How to build batched algorithm for bandit problems with general reward structures remains challenging and open.

METHODS

We propose neural network-based bandit algorithms under the batched settings. Our algorithms compute an upper confidence bound of the reward guided by the neural networks for the exploration of the contexts. Our algorithms work for the general reward setting, helped by the representation power of neural works. Meanwhile, they enjoy smaller number of policy updates due to the nature of batched algorithms.

RESULTS

The effectiveness of our proposed algorithms is demonstrated on both their theoretical guarantee and their empirical performance. From the theoretical aspect, our algorithms enjoy the same order of regret compared with their fully sequential version, while reducing the number of policy switches. From the empirical aspect, our algorithms achieve nearly the same performance as the fully sequential baseline version on simulation and several real-world datasets.

SIGNIFICANCE

Our result shows that the batched neural bandit algorithms are able to achieve nearly the same performance as the fully sequential ones, with a lower computational complexity. The result as well as the proposed algorithms have the potential to guide the future study in the related areas, such as computationally efficient algorithms in deep reinforcement learning.

Keywords: Bandits, Reinforcement learning, Deep learning

Authors' addresses: Quanquan Gu, University of California, Los Angeles, USA, qgu@cs.ucla.edu; Amin Karbasi, Yale University, USA, amin.karbasi@yale.edu; Khashayar Khosravi, Google Research NYC, USA, khosravi@google.com; Vahab Mirrokni, Google Research NYC, USA, mirrokni@google.com; Dongruo Zhou, University of California, Los Angeles, USA, drzhou@cs.ucla.edu.

Research Article