Formatting Instructions for CoRL 2020

Anonymous Author(s)
Affiliation
Address
email

Abstract: Recent success in offline Reinforcement Learning (RL) is highlighted by its adaptability to novel scenarios. One of the key reasons behind this success is the readily available nature of behavior transitions. Practical applications, on the other hand, consist of behavior policy as a sequence of demonstrations rather than a dataset of transitions. This allows one to rethink *behavior initialization* through the lens of imitation. We steer research towards this direction by answering the central question of *how can demonstrations be utilized for behavior initialization?* Our study aims to combine Imitation Learning (IL) as the behavioral aspect for offline RL. We aim to explore theoretical properties of IL-RL combinations and empirically evaluate them in light of data efficiency on a suite of locomotion and manipulation tasks. The study additionally aims to throw light on the various tradeoffs between data collection and optimal behavior and why striking a balance between the two is essential from a practical standpoint. We hope that our work serves as a motivating example for application of offline RL to practical problems.

Keywords: CoRL, Robots, Learning

1 Problem

Consider the scenario wherein a human learns to drive a car. The driver observes a teacher driving the car. This involves paying attention to crucial insights of controlling the vehicle such as steering while making a turn, accelerating during a green light and monitoring mirrors during brakes. Increasing amount of observations by the driver result in learning finer details of the task which are not available *apriori*. For instance, the driver may never learn to make a U-turn if the teacher never encountered a U-turn crossing.

Offline RL addresses this intuitive gap in learning by equiping an agent (the *driver* in above example) with the ability to stitch together portions of observations by making use of a dataset of transitions. For instance, the driver may learn to make a U-turn on its own if it observes the teacher making sharp turns and slowing down the vehicle at intersections. Adoption of transitions in the offline setting allows the agent to efficaciously tackle distributional shift between the teacher's policy and the agent's policy. To this end, it is reasonable to ask the question *how would the agent learn to make a U-turn if it never saw the teacher make sharp turns or slow down the vehicle?* More specifically, how does the agent stitch together portions of transitions with limited data?

Various scenarios make data collection imperative in the face of uncertainty. Lack of optimal behavior transitions observed by an offline RL agent may cripple its policy and result in suboptimal convergence. This allows one to rethink offline RL as an abstraction of *behavior initialization* and *learning* problems. In the first stage, the agent desires a suitable initialization point (a behavior policy or static dataset of transitions) which would serve as a guididng principle for agent's policy. The second stage comprises of learning optimal behaviors based on initialization point. The direct dependence of offline mechanims on behavior initialization highlights its pivotal role in the learning pipeline.

Submitted to the 4th Conference on Robot Learning (CoRL 2020). Do not distribute.

Modern offline RL methods resort to a black-box dataset of transitions as the initialization point. This often limits optimal behavior at the cost of data collection (as observed in *driving* example). A suitable alternative to address this limitation is by utilizing demonstrations as initialization point for the agent's policy. Similar to static transitions in offline RL, expert demonstrations provide a guiding mechanism for learner's policy in the IL setup. A suitable combination of IL-RL which trades off data collection with optimal behavior forms the center of our study.

References