Contribution ID: **64** Type: **not specified**

Model-Free Robust ϕ-Divergence Reinforcement Learning Using Both Offline and Online Data

Wednesday, June 19, 2024 1:30 PM (30 minutes)

The robust ϕ-regularized Markov Decision Process (RRMDP) framework focuses on designing control policies that are robust against parameter uncertainties due to mismatches between the simulator (nominal) model and real-world settings. This work makes \emph{two} important contributions. First, we propose a \textit{model-free} algorithm called \textit{Robust ϕ-regularized fitted Q-iteration} (RPQ) for learning an ϵ-optimal robust policy that uses only the historical data collected by rolling out a behavior policy (with \textit{robust exploratory} requirement) on the nominal model. To the best of our knowledge, we provide the \textit{first} unified analysis for a class of ϕ-divergences achieving robust optimal policies in high-dimensional systems with general function approximation. Second, we introduce the \textit{hybrid robust ϕ-regularized reinforcement learning} framework to learn an optimal robust policy using both historical data and online sampling. Towards this framework, we propose a model-free algorithm called \textit{Hybrid robust Total-variation-regularized Q-iteration} (HyTQ). Finally, we provide theoretical guarantees on the performance of the learned policies of our algorithms on systems with arbitrary large state space using function approximation.

Primary author: PANAGANTI, Kishan (California Institute of Technology)

Co-authors: WIERMAN, Adam (California Institute of Technology); MAZUMDAR, Eric (California Institute of Technology)

Presenter: PANAGANTI, Kishan (California Institute of Technology)

Session Classification: Parallel session: Online learning