Model-Based Reinforcement Learning with Value-Targeted Regression

Parameteric models allow the scalability of Reinforcement Learning (RL) to large state and action spaces. The work proposes a novel algorithm for model parameter estimation. Th transition model is assumed to admit linear parameterization. Based on this formulation, the proposed algorithm carried out model parameter estimation by recursively solving a regression problem with target as the latest value estimate. Value-targeted regression yields an upper bound on the regret $\mathcal{O}(d\sqrt{H^3T})$. The regret bound is independent of the total number of states and actions and close to the proposed lower bound $\Omega(\sqrt{HdT})$.