Review-13

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 $\Omega(\sqrt{HdT})$.

Conventional model-based RL methods explicitly estimate transition probabilities and operate on raw states. To that end, the work aims to deviate from this notion and estimates model parameters by setting up a regression problem based on the value function. Targets in regression updates are latest value estimates. Value-targeted formulation yields one-dimensional targets which eliminates the need for multivariate tuning. Addtionally, the model parameters θ are updated through a simple recursive formula. Computation is carried out by constructing upper confidence estimates of Q-values which yield value estimates. Estimated value functions are used as targets in regression with $X_{h,k}^T.\theta$ being the predicted value consisting of approximate Monte-Carlo value estimates $X_{h,k} = \mathbb{E}[V_{h+1,k}(s)|s_h^k, a_h^k]$. The empirical loss function $(X_{h,k}^T.\theta - y_{h,k})^2$ consists of the taget $y_{h,k} = V_{h+1,k}(s_{h+1}^k)$ as the latest target estimates. Loss is updated using a ridge-regression setting with a regularization term. Recursive computations involve the utility of inner product $X_{h,k}.X_{h,k}^T$.