Mach Learn (2013) 91:325–349
DOI 10.1007/s10994-013-5368-1

Minimax PAC bounds on the sample complexity of reinforcement learning with a generative model

Mohammad Gheshlaghi Azar · Rémi Munos · Hilbert J. Kappen

Received: 30 July 2012 / Accepted: 24 April 2013 / Published online: 14 May 2013

Abstract We consider the problems of learning the optimal action-value function and the optimal policy in discounted-reward Markov decision processes (MDPs). We prove new PAC bounds on the sample-complexity of two well-known model-based reinforcement learning (RL) algorithms in the presence of a generative model of the MDP: value iteration and policy iteration. The first result indicates that for an MDP with N state-action pairs and the discount factor $\gamma \in [0, 1)$ only $O(N \log(N/\delta)/((1-\gamma)^3\varepsilon^2))$ state-transition samples are required to find an ε-optimal estimation of the action-value function with the probability (w.p.) $1-\delta$. Further, we prove that, for small values of ε, an order of $O(N \log(N/\delta)/((1-\gamma)^3\varepsilon^2))$ samples is required to find an ε-optimal policy w.p. $1-\delta$. We also prove a matching lower bound of $\Theta(N \log(N/\delta)/((1-\gamma)^3\varepsilon^2))$ on the sample complexity of estimating the optimal action-value function with ε accuracy. To the best of our knowledge, this is the first minimax result on the sample complexity of RL: the upper bounds match the lower bound in terms of N, ε, δ and $1/(1-\gamma)$ up to a constant factor. Also, both our lower bound and upper bound improve on the state-of-the-art in terms of their dependence on $1/(1-\gamma)$.

Keywords Sample complexity · Markov decision processes · Reinforcement learning · Learning theory

Editor: Csaba Szepesvari.

M. Gheshlaghi Azar (✉) · H.J. Kappen
Department of Biophysics, Radboud University Nijmegen, 6525 EZ Nijmegen, The Netherlands
e-mail: m.azar@science.ru.nl

H.J. Kappen
e-mail: b.kappen@science.ru.nl

Present address:
M. Gheshlaghi Azar
School of Computer Science, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA

R. Munos
INRIA Lille, SequeL Project, 40 avenue Halley, 59650 Villeneuve d'Ascq, France
e-mail: remi.munos@inria.fr