Les mathématiques de l'apprentissage machine
« Atelier sur les défis modernes en théorie d'apprentissage »
23 au 26 avril 2018

Mathematics of Machine Learning
"Workshop on Modern Challenges of Learning Theory"
April 23-26, 2018

A unified view of entropy-regularized Markov decision processes

Gergely Neu *

gergely.neu@gmail.com

Entropy regularization, while a standard technique in the online learning toolbox, has only been recently discovered by the reinforcement learning community: In recent years, numerous new reinforcement learning algorithms have been derived using this principle, largely independently of each other. So far, a general framework for these algorithms has remained elusive. In this work, we propose such a general framework for entropy-regularized average-reward reinforcement learning in Markov decision processes (MDPs). Our approach is based on extending the linear-programming formulation of policy optimization in MDPs to accommodate convex regularization functions. Our key result is showing that using the conditional entropy of the joint state-action distributions as regularization yields a dual optimization problem closely resembling the Bellman optimality equations. This result enables us to formalize a number of state-of-the-art entropy-regularized reinforcement learning algorithms as approximate variants of Mirror Descent or Dual Averaging, and thus to argue about the convergence properties of these methods. In particular, we show that the exact version of the TRPO algorithm of Schulman et al. (2015) actually converges to the optimal policy, while the entropy-regularized policy gradient methods of Mnih et al. (2016) may fail to converge to a fixed point.

*DTIC, Universitat Pompeu Fabra, C/ Roc Boronat 138, 08018 Barcelona, SPAIN