Email Address:

zp2130@caa.columbia.edu (mailto:zp2130@caa.columbia.edu)

p@caa.columbia.edu (mailto:p@caa.columbia.edu)

Randomized Linear Programming Solves the Discounted Markov Decision Problem In Nearly-Linear (Sometimes Sublinear) Run Time

Randomized Linear Programming Solves the Discounted Markov Decision Problem In Nearly-Linear (Sometimes Sublinear) Run Time

Randomized Linear Programming Solves the Discounted Markov Decision Problem In Nearly-Linear (Sometimes Sublinear) Run Time (https://blogs.cuit.columbia.edu/zp2130/files/2019/07/Randomized-Linear-Programming-Solves-the-Discounted-Markov-Decision-Problem-In-Nearly-Linear-Sometimes-Sublinear-Run-Time.pdf)

The nonlinear Bellman equation = $|S| \times (|S|\,|A|)$ linear programming problem: Primal-Dual LP

Primal LP

$$min.\,(1-\gamma)\,\mathbf{q}^T\mathbf{v} \qquad s.t.\,(I-\gamma P_a)\,\mathbf{v}-\mathbf{r}_a \geq 0, \forall a \in A, \quad (1)$$
$$\forall i \in S \qquad r_a(i) = \sum_{j \in S} p_{ij}(a) r_{ij}(a)$$

Dual LP

$$max. \sum_{a \in A} \mu_a^T \mathbf{r}_a \qquad s.t. \sum_{a \in A} \left(I-\gamma P_a^T\right)\mu_a = (1-\gamma)\,\mathbf{q}, \mu_a \geq 0, \forall a \in A. \quad (2)$$

Minmax Problem

$$min_{\mathbf{v} \in V}\; max_{\mu \in U_{\theta,\mathbf{q}}}\,(1-\gamma)\,\mathbf{q}^T\mathbf{v} + \sum_{a \in A} \mu_a^T\left((\gamma P_a - I)\,\mathbf{v} + \mathbf{r}_a\right) \quad (3)$$

$$V = \left\{ \|\mathbf{v}\|_\infty \leq \frac{1}{1-\gamma}, \mathbf{v} \geq 0 \right\}, U_{\theta,\mathbf{q}} = \left\{ \mathbf{e}^T \mu = 1, \mu \geq 0, \sum_{a \in A} \mu_a \geq \theta \mathbf{q} \right\}$$

edit (https://blogs.cuit.columbia.edu/zp2130/wp-admin/post.php?post=5425&action=edit)

Author: Z Pei (https://blogs.cuit.columbia.edu/zp2130/author/zp2130/) on July 15, 2019

Categories: AI (https://blogs.cuit.columbia.edu/zp2130/category/ai/), Algorithm (https://blogs.cuit.columbia.edu/zp2130/category/algorithm/), Computational Complexity (https://blogs.cuit.columbia.edu/zp2130/category/computational-complexity/), Control (https://blogs.cuit.columbia.edu/zp2130/category/control/), Convergence Rate (https://blogs.cuit.columbia.edu/zp2130/category/convergence-rate/), Machine Learning (https://blogs.cuit.columbia.edu/zp2130/category/machine-learning/), Reinforcement Learning (https://blogs.cuit.columbia.edu/zp2130/category/reinforcement-learning/), RL (https://blogs.cuit.columbia.edu/zp2130/category/rl/)

Tags: AI (https://blogs.cuit.columbia.edu/zp2130/tag/ai/), Algorithm (https://blogs.cuit.columbia.edu/zp2130/tag/algorithm/), Computational Complexity (https://blogs.cuit.columbia.edu/zp2130/tag/computational-complexity/), Control (https://blogs.cuit.columbia.edu/zp2130/tag/control/), Convergence Rate (https://blogs.cuit.columbia.edu/zp2130/tag/convergence-rate/), Machine Learning (https://blogs.cuit.columbia.edu/zp2130/tag/machine-learning/), Reinforcement Learning (https://blogs.cuit.columbia.edu/zp2130/tag/reinforcement-learning/), RL (https://blogs.cuit.columbia.edu/zp2130/tag/rl/)

Other posts

KL Divergence (https://blogs.cuit.columbia.edu/zp2130/kl_divergence/) «» Solving H-horizon, Stationary Markov Decision Problems In Time Proportional To Log(H) (https://blogs.cuit.columbia.edu/zp2130/paul_tseng_1990/)

Last posts

- Symbolic Netlist to Innovus-friendly Netlist (https://blogs.cuit.columbia.edu/zp2130/symbolic_netlist_to_innovus-friendly_netlist/)
- Finite-Sample Convergence Rates for Q-Learning and Indirect Algorithms (https://blogs.cuit.columbia.edu/zp2130/finite-sample_convergence_rates_for_q-learning_and_indirect_algorithms/)
- Solving H-horizon, Stationary Markov Decision Problems In Time Proportional To Log(H) (https://blogs.cuit.columbia.edu/zp2130/paul_tseng_1990/)
- Randomized Linear Programming Solves the Discounted Markov Decision Problem In Nearly-Linear (Sometimes Sublinear) Run Time (https://blogs.cuit.columbia.edu/zp2130/randomized_linear_programming_solves_the_discounted_markov_decision_problem_in_nearly-linear_sometimes_sublinear_run_time/)
- KL Divergence (https://blogs.cuit.columbia.edu/zp2130/kl_divergence/)
- The Asymptotic Convergence-Rate of Q-learning (https://blogs.cuit.columbia.edu/zp2130/the_asymptotic_convergence-rate_of_q-learning/)
- Hierarchical Apprenticeship Learning, with Application to Quadruped Locomotion (https://blogs.cuit.columbia.edu/zp2130/hierarchical_apprenticeship_learning_with_application_to_quadruped_locomotion/)
- Policy Gradient Methods (https://blogs.cuit.columbia.edu/zp2130/policy_gradient_methods/)
- Actor-Critic Algorithms for Hierarchical Markov Decision Processes (https://blogs.cuit.columbia.edu/zp2130/actor-critic_algorithms_for_hierarchical_markov_decision_processes/)
- Hierarchical Deep Reinforcement Learning: Integrating Temporal Abstraction and Intrinsic Motivation (https://blogs.cuit.columbia.edu/zp2130/hierarchical_deep_reinforcement_learning_integrating_temporal_abstraction_and_intrinsic_motivation/)