

Reinforcement Learning with Function Approximation—Done Right



Gugan Thoppe
Indian Institute of Science

Tuesday, July 22
1:30 PM • MSEE 239

Abstract

Policy Iteration (PI) is one of the cornerstones of Reinforcement Learning (RL), which has and continues to inspire several modern RL algorithms. PI's fame rests on its guarantees of monotonic policy improvement and convergence to the optimal policy. However, in practice, approximate PI variants are necessitated to handle the large state and action spaces. Despite decades of research, variants that preserve the classic PI guarantees have remained elusive, even under simple, linear function approximation. In this talk, I will introduce Reliable Policy Iteration (RPI) that addresses this issue. It replaces the common projection and Bellman minimization during policy evaluation with a Bellman-based constrained optimization. We prove that not only does RPI confer textbook monotonicity on its value estimates, but these estimates also lower bound the true return. Also, their limit partially satisfies the unprojected Bellman equation, emphasizing RPI's natural fit within RL. RPI is the first algorithm with such monotonicity and convergence guarantees under function approximation. For practical use, we provide a model-free variant of RPI that amounts to a novel critic. It can be readily integrated into primary model-free PI implementations such as DQN and DDPG. In classical control tasks, such RPI-enhanced variants consistently maintain their lower-bound guarantee while matching or surpassing the performance of all baseline methods.

This talk draws on our recent preprint, “Reliable Critics: Monotonic Improvement and Convergence Guarantees for Reinforcement Learning,” co-authored with Eshwar S. R., Aditya Gopalan, and Gal Dalal (<https://arxiv.org/pdf/2506.07134>).

Bio

Gugan Thoppe is an Assistant Professor in the Computer Science and Automation (CSA) department at the Indian Institute of Science since 2019. He received his Ph.D. in 2016 from the Tata Institute of Fundamental Research (TIFR), Mumbai. He then completed postdoctoral research at Technion Institute of Technology, Israel (2015-17) and Duke University, USA (2017-19). His research is supported by CEFIPRA, Walmart, NPCI, and DST-SERB. He has won the Pratiksha Trust's Young Investigator Award, the IISc Award for Excellence in Teaching, and the TIFR Award for the best Ph.D. thesis. He is also part of ACM India's Eminent Speaker Panel. His research interests include reinforcement learning, distributed learning, and stochastic approximation.

Host

Vijay Gupta, gupta869@purdue.edu