Pulkit Agrawal

Title: Dilemmas in Reinforcement and Imitation Learning (edited)

Abstract: A reinforcement learning agent faces the fundamental exploration-exploitation dilemma: should it try out more actions in the hope of finding higher rewards (i.e., exploration), or should it continue taking actions with the best performance until now (i.e., exploitation)? In a different context, if the agent can access a teacher, it faces another dilemma: whether to mimic the teacher's action or learn from rewards. On the one hand, imitating a teacher can simplify exploration that improves performance, but on the other hand, if the teacher is sup-optimal, imitating can limit the agent's performance. Further, in some cases, if the teacher is too good, it's only possible to mimic by taking different actions that provide the information necessary for imitation, posing another dilemma. I'll talk about our recent progress in automatically (but approximately) balancing these dilemmas without hyperparameter tuning. The result is practical algorithms with superior exploration, more efficient solutions to POMDPs, and the promise to reduce dependence on reward design. Lastly, I will contextualize these advances with the problem of building robotic controllers for dynamic and contact-rich problems of dexterous manipulation and agile locomotion.