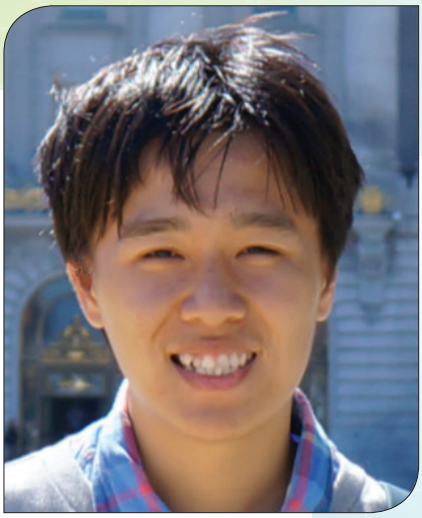
COLLEGE OF ENGINEERING
COMPUTER SCIENCE & ENGINEERING
UNIVERSITY OF MICHIGAN

CSE Dissertation Defense



Nan Jiang

A Theory of Model Selection in Reinforcement Learning

ABSTRACT: Reinforcement Learning (RL) is a machine learning paradigm where an agent learns to accomplish sequential decision-making tasks from experience. Most empirical successes of RL today are restricted to simulated environments, where hyperparameters are tuned by trial and error using unlimited data. In contrast, collecting data with active intervention and control in the real world can be costly, time-consuming, and sometimes unsafe. Choosing the hyperparameters and understanding their effects in face of these data limitations, i.e., model selection, is an open direction that is crucial to the real-world applications of RL.

In this thesis, I present theoretical results that improve our understanding of 3 hyperparameters: planning horizon, state representation (abstraction), and reward function. The 1st part focuses on the interplay between planning horizon and limited amount of data, and establishes a formal explanation for how a long planning horizon can cause overfitting. The 2nd part considers the problem of choosing the right state abstraction using limited batch data; I show that cross-validation type methods require importance sampling and suffer from exponential variance, and a novel regularization-based algorithm enjoys an oracle-like property. The 3rd part studies reward misspecification and tries to resolve it by leveraging expert demonstrations, which is inspired by AI safety concerns and is closely related to inverse RL.

A recurring theme of the thesis is the deployment of formulations and techniques from other machine learning theory (mostly statistical learning theory): the planning horizon work explains the overfitting phenomenon by making a formal analogy to empirical risk minimization and proving planning loss bounds that are similar to generalization error bounds; the main result in the abstraction selection work takes the form of an oracle inequality, which is a concept from structural risk minimization for model selection in supervised learning; the inverse RL work provides a mistake-bound type analysis under arbitrarily chosen environments, which can be viewed as a form of no-regret learning. Overall, by borrowing ideas from mature theories of machine learning, we can develop analogies for RL that allow us to better understand the impact of hyperparameters, and develop algorithms that automatically set them in an effective manner.

Chair: Prof. Satinder Singh Baveja

Wednesday, May 10, 2017 1:00 – 3:00 pm 3725 Beyster Building