Model Free Control

Introduction

Marius Lindauer

Leibniz Universität Hannover

Automated Machine Learning Hannover

Overview

- Last time: Policy evaluation with no knowledge of how the world works
  - Aim: We wanted to know how well a given policy would perform
  - MDP model (e.g., transition function and reward function) not given

Overview

- Last time: Policy evaluation with no knowledge of how the world works
  - Aim: We wanted to know how well a given policy would perform
  - MDP model (e.g., transition function and reward function) not given
- This time: Control (making decisions) without a model of how the world works
  - We have to search for a well-performing policy
  - We still don't know the MDP model
  - We assume that we can model everything by table look-ups

Recall: Reinforcement Learning involves

- Optimization
- Delayed consequences
- Exploration
- Generalization

Learning to Control Involves

- Optimization: Goal is to identify a policy with high expected rewards (similar to before on computing an optimal policy given an MDP)
- Delayed consequences: May take many time steps to evaluate whether an earlier decision was good or not
- Exploration: Necessary to try different actions to learn what actions can lead to high rewards
- (Generalization – deferred to later)

Model-free Control Examples

- Many applications can be modeled as an MDP: Backgammon, Go, Robot locomotion, Helicopter flight, Robocup soccer, Autonomous driving, Customer ad selection, Invasive species management, Patient treatment
- For many of these and other problems either:
  - MDP model is unknown but can be sampled
  - MDP model is known but it is computationally infeasible to use directly, except through sampling

Automated Machine Learning Hannover

On and Off-Policy Learning

- On-policy learning
  - Direct experience
  - Learn to estimate and evaluate a policy from experience obtained from following that policy

On and Off-Policy Learning

- On-policy learning
  - Direct experience
  - Learn to estimate and evaluate a policy from experience obtained from following that policy
- Off-policy learning
  - Learn to estimate and evaluate a policy using experience gathered from following a different policy

Automated Machine Learning Hannover