CS885 Reinforcement Learning Lecture 1a: May 2, 2018

Course Introduction

[SutBar] Chapter 1, [Sze] Chapter 1

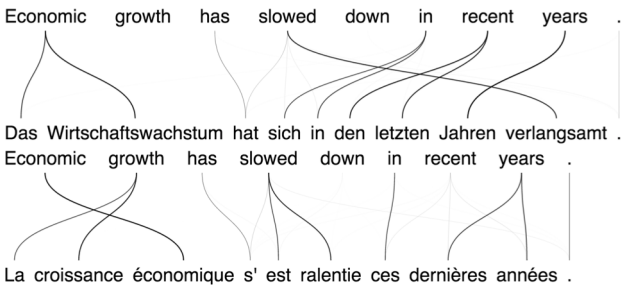
Outline

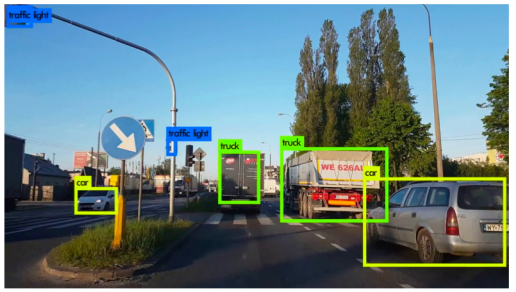
- Introduction to Reinforcement Learning
- Course website and logistics

Machine Learning

- Traditional computer science
  - Program computer for every task

- New paradigm
  - Provide examples to machine
  - Machine learns to accomplish a task based on the examples





Machine Learning

- Success mostly due to supervised learning
  - Bottleneck: need lots of labeled data
- Alternatives
  - Unsupervised learning, semi-supervised learning
  - Reinforcement Learning

What is Reinforcement Learning?

- Reinforcement learning is also known as
  - Optimal control
  - Approximate dynamic programming
  - Neuro-dynamic programming

- Wikipedia: reinforcement learning is an area of machine learning inspired by behavioural psychology, concerned with how software **agents** ought to take **actions** in an **environment** so as to maximize some notion of cumulative **reward**.

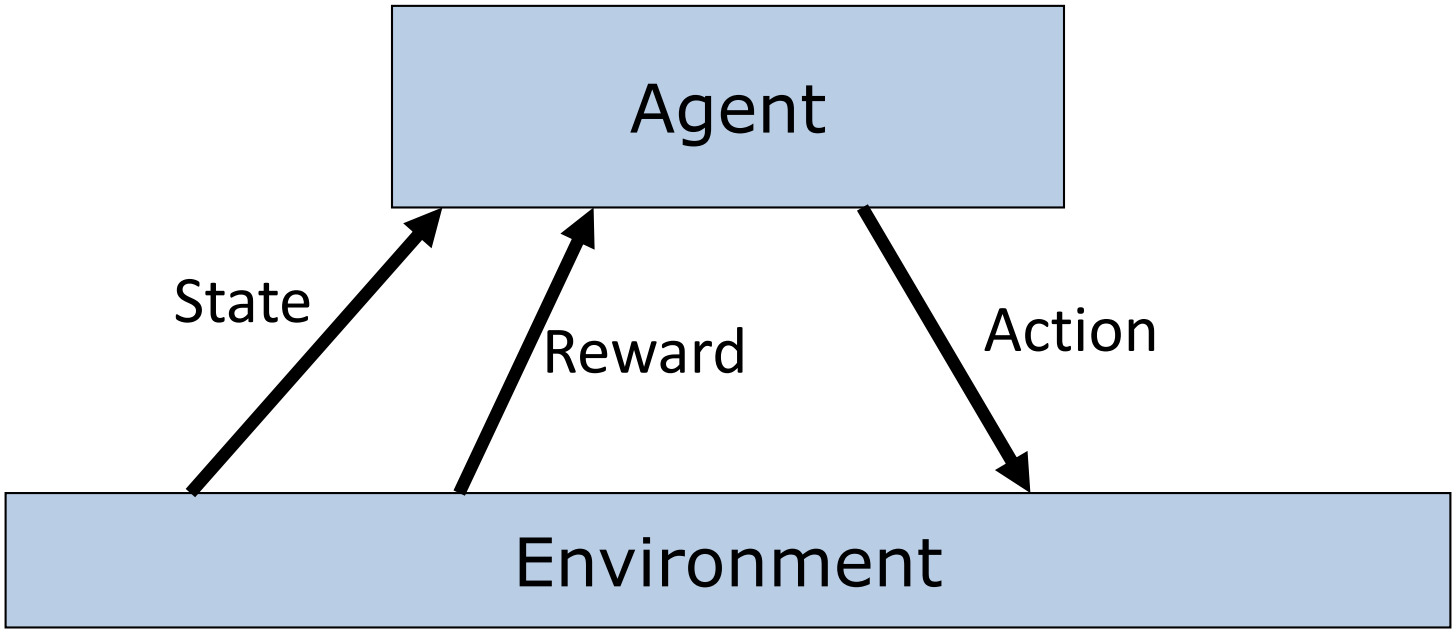
Animal Psychology

- Negative reinforcements:
  - Pain and hunger
- Positive reinforcements:
  - Pleasure and food

- Reinforcements used to train animals

- Let's do the same with computers!

Reinforcement Learning Problem



Goal: Learn to choose actions that maximize rewards

RL Examples

- Game playing (go, atari, backgammon)
- Operations research (pricing, vehicle routing)
- Elevator scheduling
- Helicopter control
- Spoken dialog systems
- Data center energy optimization
- Self-managing network systems
- Autonomous vehicles
- Computational finance

Operations research

- Example: vehicle routing

- **Agent:** vehicle routing software
- **Environment:** stochastic demand
- **State:** vehicle location, capacity and depot requests
- **Action:** vehicle route
- **Reward:** - travel costs

Robotic Control

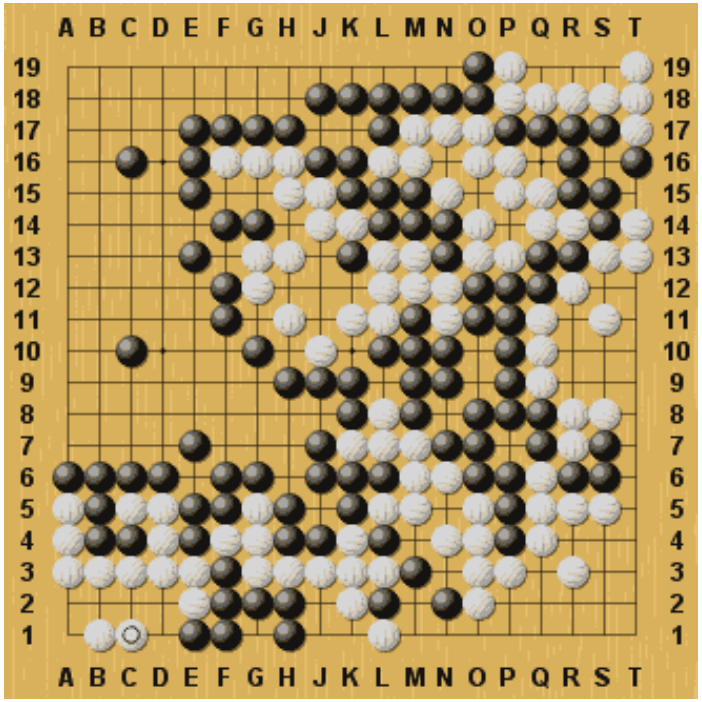
- Example: helicopter control

- **Agent:** controller
- **Environment:** helicopter
- **State:** position, orientation, velocity and angular velocity
- **Action:** collective pitch, cyclic pitch, tail rotor control
- **Reward:** - deviation from desired trajectory

- 2008 (Andrew Ng): automated helicopter wins acrobatic competition against humans

Game Playing

- Example: Go (one of the oldest and hardest board games)



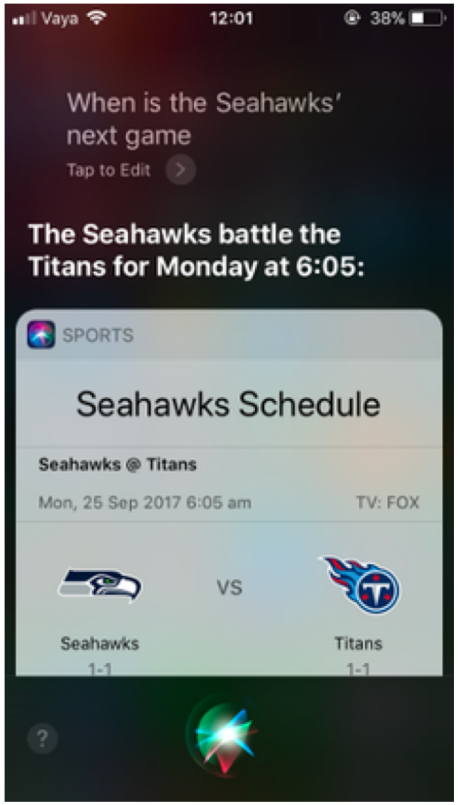
- **Agent:** player
- **Environment:** opponent
- **State:** board configuration
- **Action:** next stone location
- **Reward:** +1 win / -1 loose

- 2016: AlphaGo defeats top player Lee Sedol (4-1)
  - Game 2 move 37: AlphaGo plays unexpected move (odds 1/10,000)

Conversational agent

- **Agent:** virtual assistant
- **Environment:** user
- **State:** conversation history
- **Action:** next utterance
- **Reward:** points based on task completion, user satisfaction, etc.

- Today: active area of research



Computational Finance

- Automated trading

- **Agent:** trading software
- **Environment:** other traders
- **State:** price history
- **Action:** buy/sell/hold
- **Reward:** amount of profit

Example: how to purchase a large # of shares in a short period of time without affecting the price

Reinforcement Learning

- Comprehensive, but challenging form of machine learning
  - Stochastic environment
  - Incomplete model
  - Interdependent sequence of decisions
  - No supervision
  - Partial and delayed feedback

- **Long term goal**: lifelong machine learning