CMPUT 397 Reinforcement Learning:

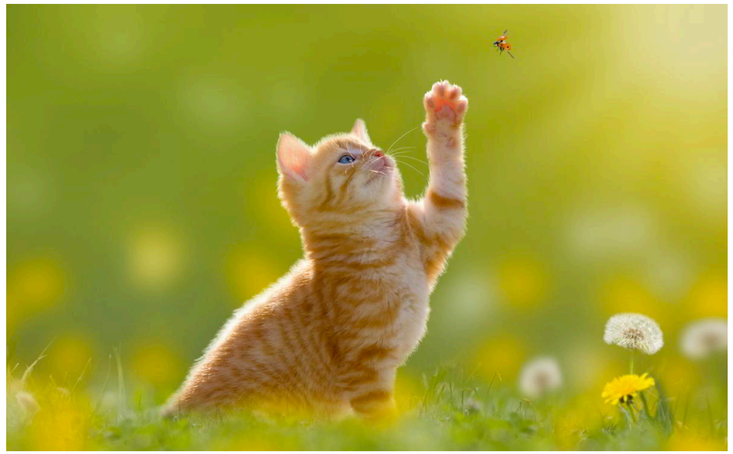
Introduction

Rupam Mahmood

January 6, 2020

Goal is a compact way of describing intelligent behaviors

A gigantic granite boulder



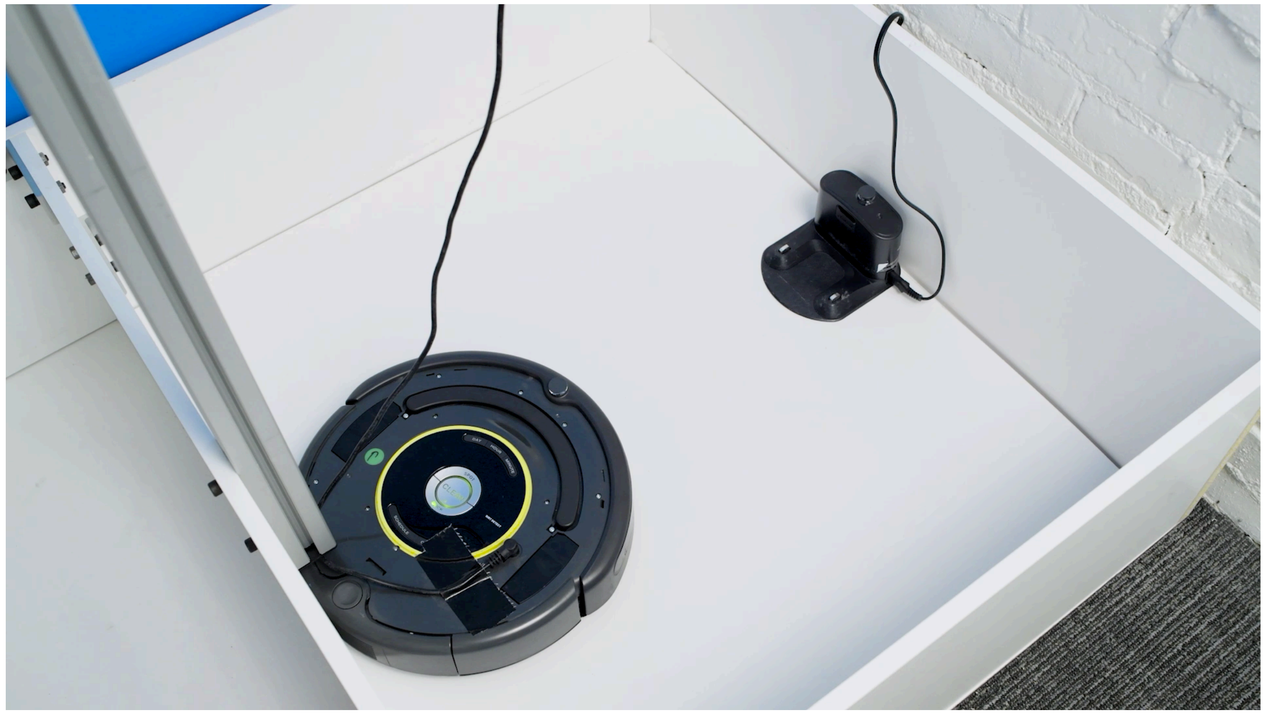
A cat catching a ladybug

(A goal-achieving system)

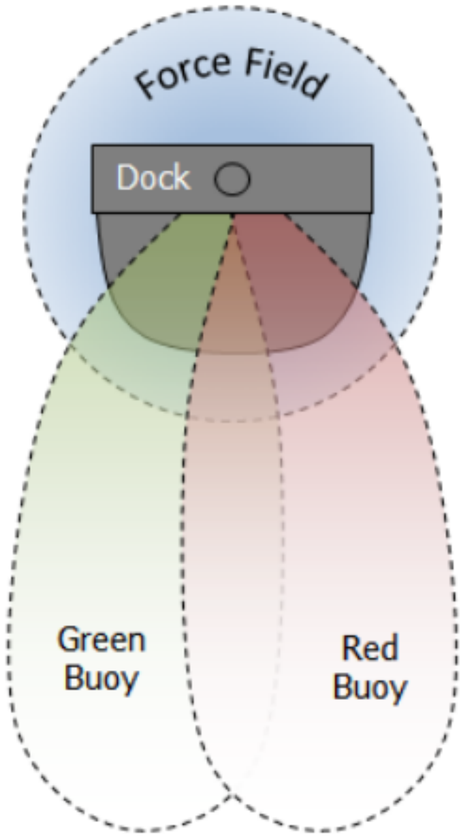
Intelligence is the computational part of the ability to achieve goals in the world

AI is the study of goal-achieving systems

Consider the goal of docking to a charging station



Dock beam configuration



AI systems are conventionally hand-engineered

If perceive *situation/state S*; then choose **action A**

Decision-making can be seen as a state-to-action mapping known as a *policy*

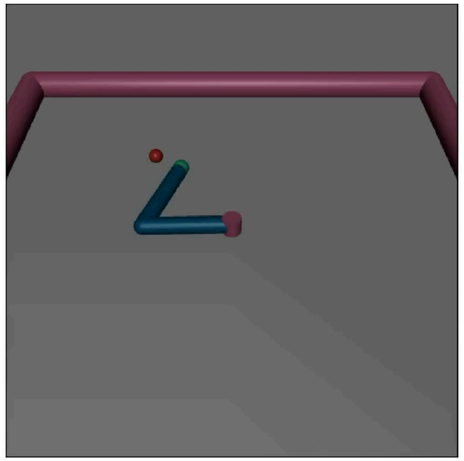
Conventionally, such maps are pre-determined, e.g., through hand-engineering

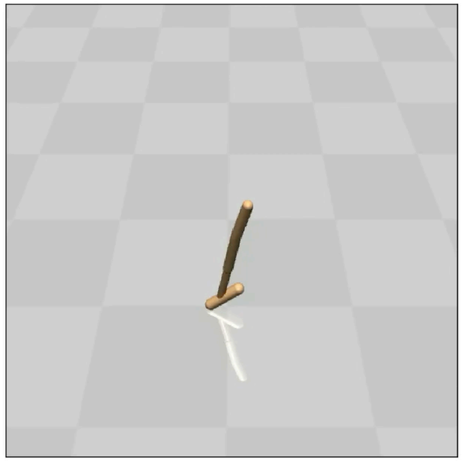
The docking behavior of a scripted agent provided by the manufacturer

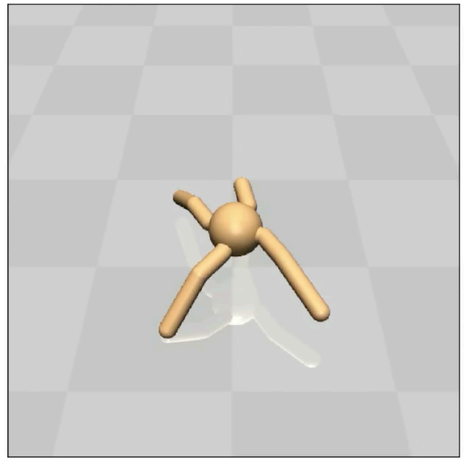
The scope for decision making is numerous.

Who should build the policy?

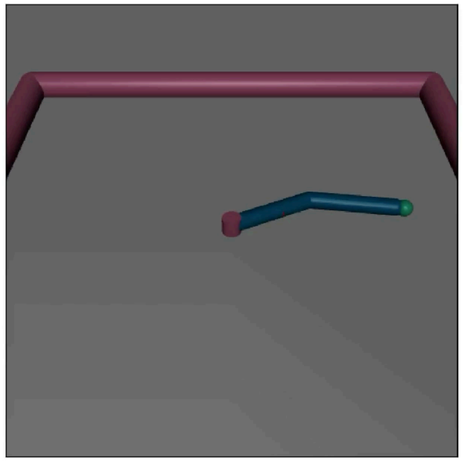
Initial behavior (random map)

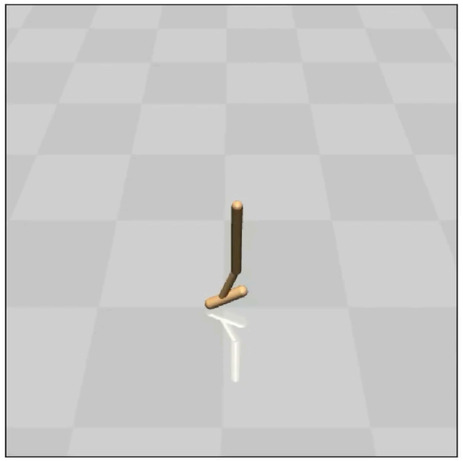


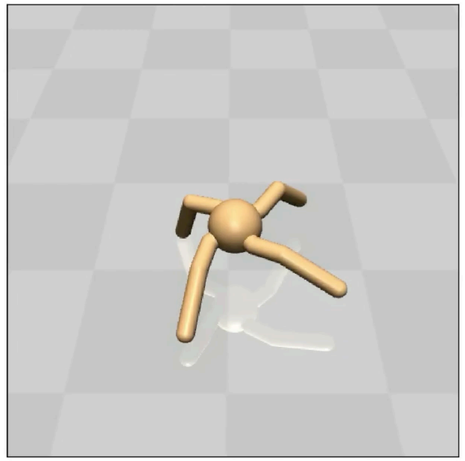




After learning the policy



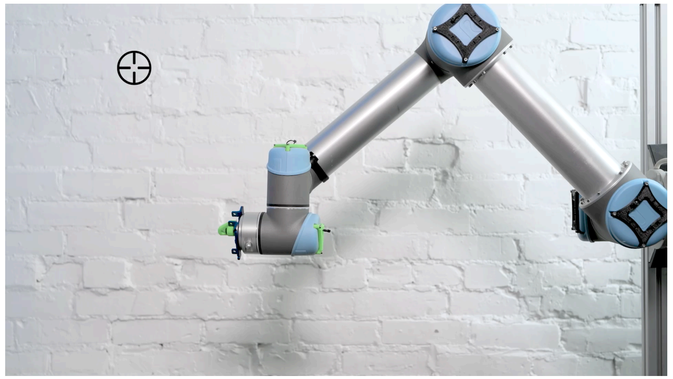




Same mechanism used to learn the policy in all three cases

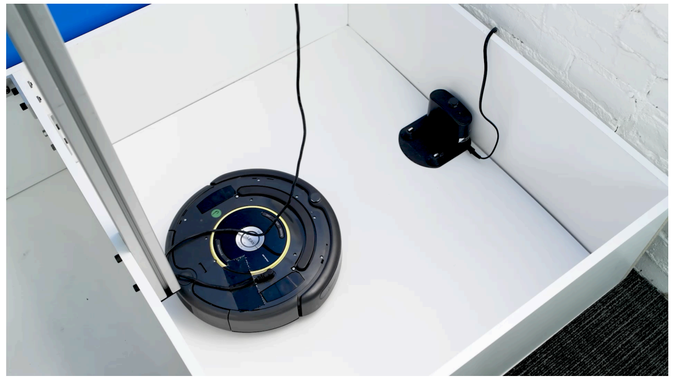
At Kindred, we developed a robot learning platform to show the generality of RL systems using physical robots

Initial behavior

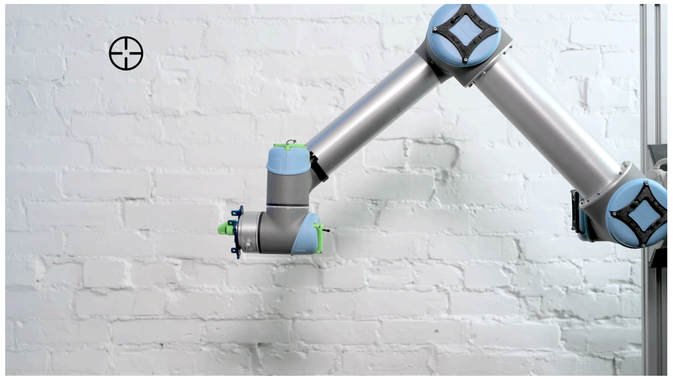


Initial behavior

Initial behavior

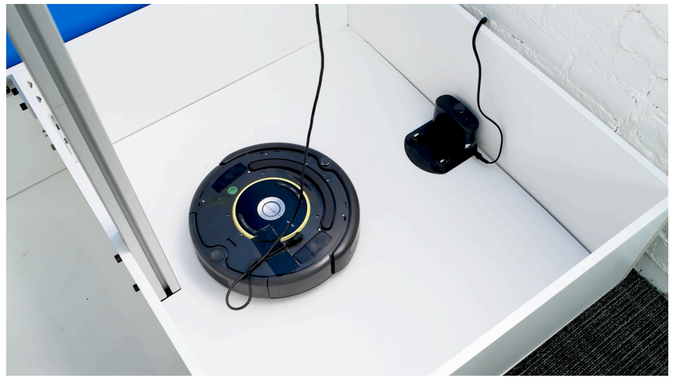


Learned behavior



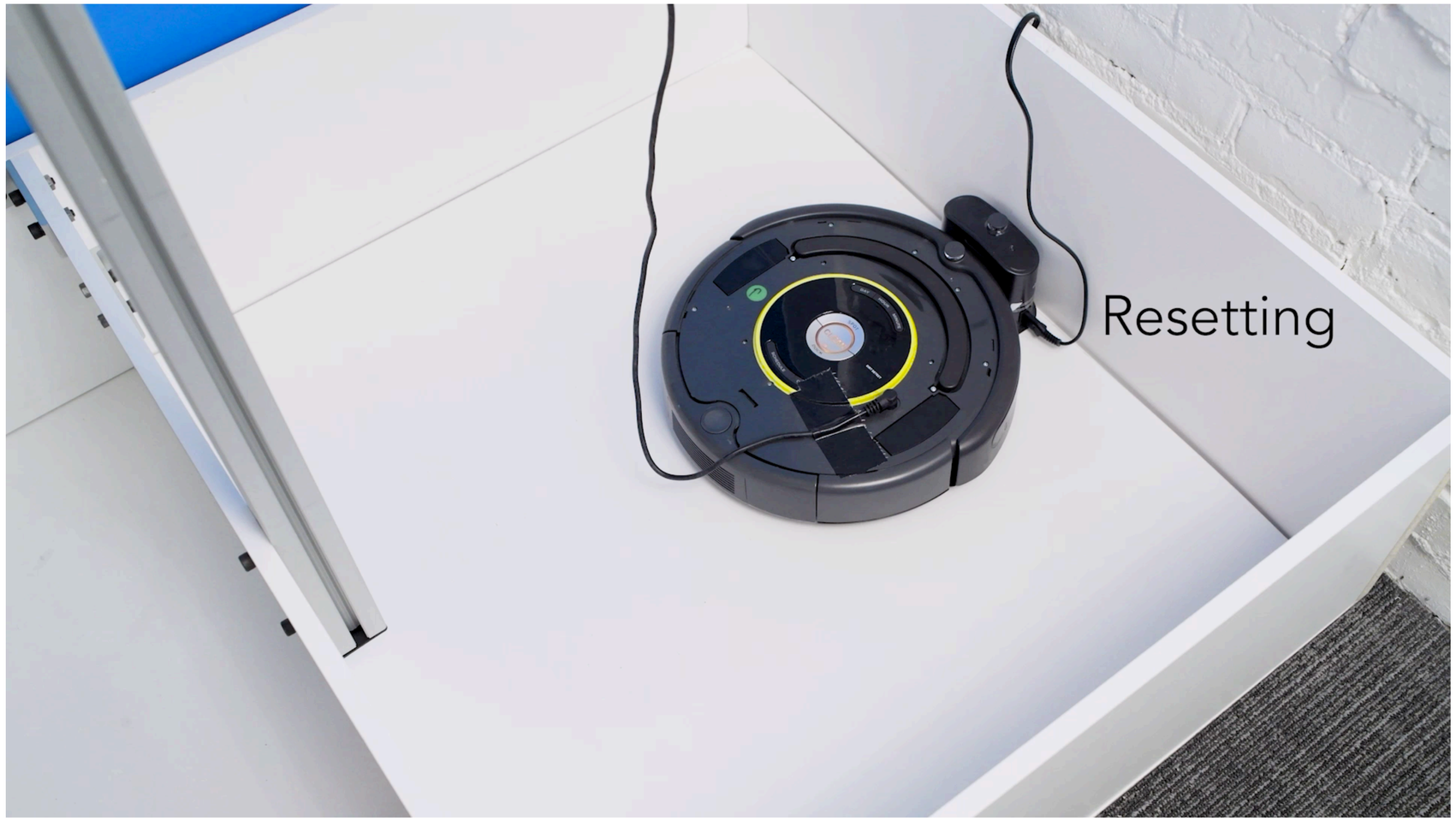
Learned behavior

Learned behavior



https://github.com/kindredresearch/SenseAct

The docking behavior of a learned policy



Learning to dock reproduced by students in CMPUT 652: RL with Robots (Fall 2019)

Archit Sakhadeo, Parash Rahman, Hager Radi, Mohamed Elsayed

This Course

- ✓ In this course, we will develop the conceptual understanding that builds up to algorithms for prediction and control learning
  e.g., temporal-difference learning, Q learning
- ✓ We will learn about the technical frameworks for these algorithms
  e.g., Bandits and Markov Decision Processes
- ✓ A clear understanding of probabilities and expectations is required

Instruction Team

- ✓ Rupam Mahmood (Faculty @UofA, RLAI, Amii, CIFAR, Kindred)
- ✓ TAs (graduate students doing RL research):
  - Banafsheh Rafiee
  - Xutong Zhao
  - Shivam Garg

Contacting Us

- ✓ Use the course discussion feature on eClass
- ✓ Meet with the instruction team during office hours

Prerequisites

- ✓ Some comfort or interest in thinking abstractly and with mathematics
- ✓ Elementary statistics, probability theory
  - conditional expectations of random variables
- ✓ Basic linear algebra: vectors, vector equations, gradients
- ✓ Programming skills (Python)

Textbook

- ✓ Readings will be from: Reinforcement Learning: An Introduction, by R Sutton and A Barto, MIT Press.
  - available freely online: http://incompleteideas.net/book/the-book-2nd.html

Course Information

- ✓ Github pages: https://armahmood.github.io/rlcourse/
  - Syllabus
  - updated schedule containing slides, assignment links & deadlines
- ✓ Course eClass page:
  - discussions, questions, announcements & course information
- ✓ Coursera RL Specialization:
  - Videos, assignments & project

Coursera RL Specialization

- ✓ We have our own private session, instruction on registering @ eClass
- ✓ RL Specialization has four courses
- ✓ Each course divided into 4-to-5 weeks
- ✓ To avoid confusions, we refer to *Coursera Courses* as **mini-courses** and *Coursera weeks* as **modules**
- ✓ The Markov Decision Processes covered in Coursera Course 1 and week 2 is referred to as **mini-course 1, module 2**

Coursera RL Specialization (continued)

- ✓ Mini-course 1: Fundamentals of RL (4 modules)
- ✓ Mini-course 2: Sample-based learning methods (4 modules: 2-5)
- ✓ Mini-course 3: Prediction and Control with Function Approximation (4 modules)
- ✓ Mini-course 4: Capstone project (4 modules)

How We Use Coursera RL Specialization

- ✓ Each week starting from next will be based on one module
- ✓ Each module has videos, 1 practice quiz, & 1 graded assignment
- ✓ Practice quiz is usually due on Sunday the week before
- ✓ Graded assignment is usually due on Friday of the week for the module
- ✓ For example, next week is K-armed bandit (mini-course 1, module 1)
- ✓ Its practice quiz due on Jan 12 (Sun), its assignment due on Jan 17 (Fri)
- ✓ Capstone modules will not run in-class but will have a deadline

Evaluation

- ✓ Practice quizzes (80% pass, 10 out of 11) 10%
- ✓ Assignments (graded quizzes/notebooks on Coursera) 30%
- ✓ Midterm exam 20%
- ✓ Final exam 30%
- ✓ Project 10%

More on Practice Quizzes and Assignments

- ✓ All practice quizzes and assignments have five attempts
- ✓ The surest way to do well in practice quizzes is to
  - watch the lecture videos (at most 1 hour of time)
  - complete the readings
- ✓ Assignments are graded quizzes or notebooks

Grading

- ✓ Grades are not based on a curve
- ✓ You can see your approximate ranking in eClass throughout the course
- ✓ Letter grades are provided by ranges of percentages

Collaboration

- ✓ Working together to solve the problems is encouraged
- ✓ But you must write-up your answers individually
- ✓ Do not take written notes during discussions
- ✓ You must acknowledge all the people you talked with in solving the problems

Academic Integrity

- The University of Alberta is committed to the highest standards of academic integrity and honesty. Students are expected to be familiar with these standards regarding academic honesty and to uphold the policies of the University in this respect. Students are particularly urged to familiarize themselves with the provisions of the Code of Student Behavior (online at www.ualberta.ca/secretariat/appeals.htm) and avoid any behavior which could potentially result in suspicions of cheating, plagiarism, misrepresentation of facts and/or participation in an offence. Academic dishonesty is a serious offence and can result in suspension or expulsion from the University.

AI Seminars

- ✓ Typically Friday noons @ CSC 3-33
- ✓ Free Pizzas
- ✓ Great AI topics and speakers