Artificial Intelligence (CS303)

Lecture 10: Learning Principles

Hints for this lecture

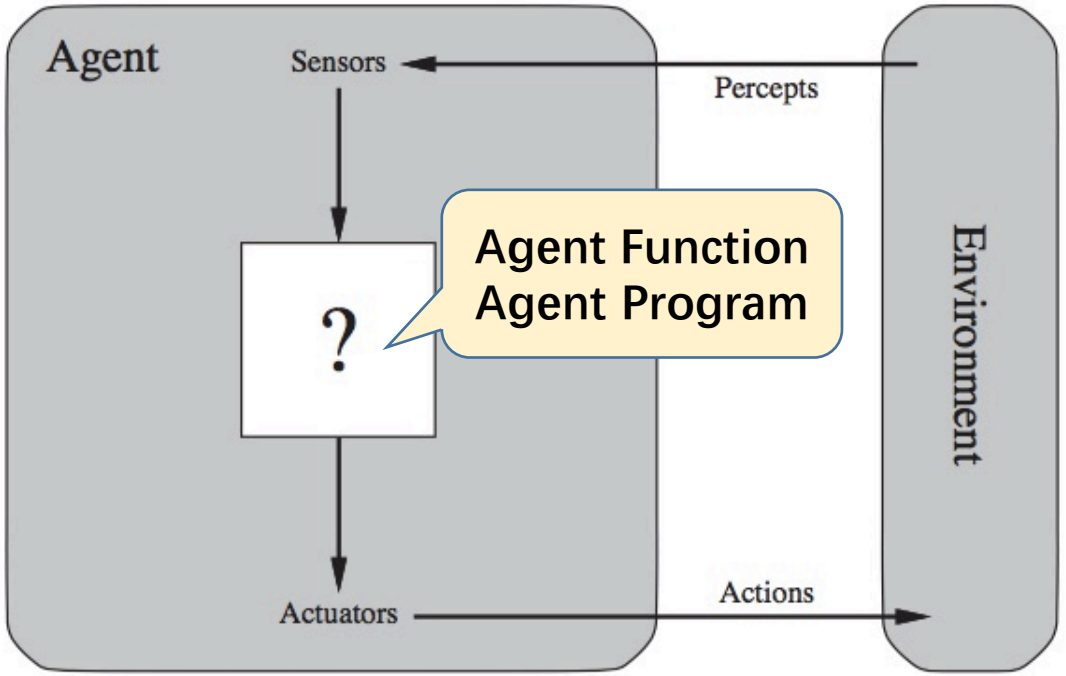
- **Learning = Search for hypothesis/functions that generalize well.**

Outline of this lecture

- **What is Learning**
- **Key Questions for Learning**
- **Learning Paradigms and Principles**

I. What is Learning

Why Learning is Important for AI?



- **When interact with the environment, we (human) take some input and give some output.**

Why Learning is Important for AI?

- Intuitively, our behavior is not determined by some rigid/static program, i.e., we might behave different as growing up.

- Learning is universal and a major source of our behavior changing.

- From an engineering perspective, it is also intractable to implement/code an agent function/program that encompasses all possible input/output pairs in the complex world. Thus an AI with learning ability would also help.

What is Learning (in AI)?

- Examples for human learning
  - Learn similarity between objects
  - Learn to recognize objects
  - Learn to pass the college entrance examination

- A process to achieve an agent function from scratch, or to improve the agent function based on observations received from the environment.

- More concretely: Given some observations (data) from the environment, and (possibly) some prior knowledge, how could an agent improve its agent function?

II. Key Questions for Learning

Key Questions for Learning

- In AI, a learning process is implemented by algorithms and programs.

- Given some observations (data) from the environment, and (possibly) some prior knowledge, how could an agent improve its agent function?
  - What is the format of the data? (data representation)
  - What is the prior knowledge?
  - What does the agent function look like? (model representation)
  - How to measure the "improvement" ? (objective function)
  - What is the learning algorithm? (to get a good agent function)

Representation + Algorithm + Evaluation

Key Questions for Learning

- In AI, a learning process is implemented by algorithms and programs.

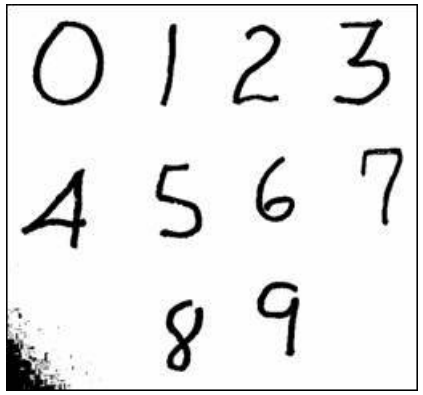
- Given some observations (data) from the environment, and (possibly) some prior knowledge, how could an agent improve its agent function?
  - What is the format of the data? (data representation)
  - What is the prior knowledge?
  - What does the agent function look like? (model representation)
  - How to measure the "improvement"? (objective function)
  - What is the learning algorithm?

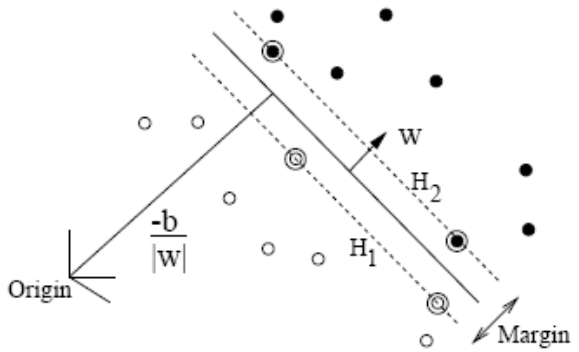
Search in a model/hypothesis space

Representation + Algorithm + Evaluation

An example

- **Hand-written Digit Recognition**
  - **Data: image of a hand-written digit**
  - **No prior knowledge**
  - **Agent function: a linear function of input**
  - **Improvement: how many images can be recognized correctly.**





III. Learning Paradigms and Principles

Inductive vs. Deductive

- **Inductive: Learning from data (prior knowledge not required)**
  - **E.g., Hand-written digit recognition**

- **Deductive: Learning from prior knowledge and data.**
  - **E.g., People in the same country speak the same language (prior knowledge) + I speak Chinese (observation/data), then you speak Chinese**

Supervised, Unsupervised and Reinforcement Learning

- **Supervised learning: the correct answer is available to the learning algorithm.**

- **Unsupervised: no correct answer is available.**

- **Reinforcement: the only feedback is the reward of an output (action), e.g., the output is correct or not (in case of incorrect, the correct answer is not given).**

Learning Principles

- Generalization: the learned agent function is expected to be able to handle previously unseen situations. （举一反三，不要刻舟求剑）
- How to estimate the generalization? : Key to design the objective function for learning.
- Occam's Razor: An explanation of data should be made as simple as possible, but no simpler, i.e., model complexity should be controlled.
- Overfitting: the model perfectly fit the seen data perfectly, but cannot generalize well.
- Occam's Razor and overfitting is closely related.

Summary

- The learning process for an AI is basically the search for a model/hypothesis/rule that can map an input from the environment to an appropriate output, such that we don’ t need to program them in advance.

- Different settings of learning may lead to different learning algorithms.

- Generalization is the ultimate goal of learning, but usually uneasy to measure/achieve.

To be continued