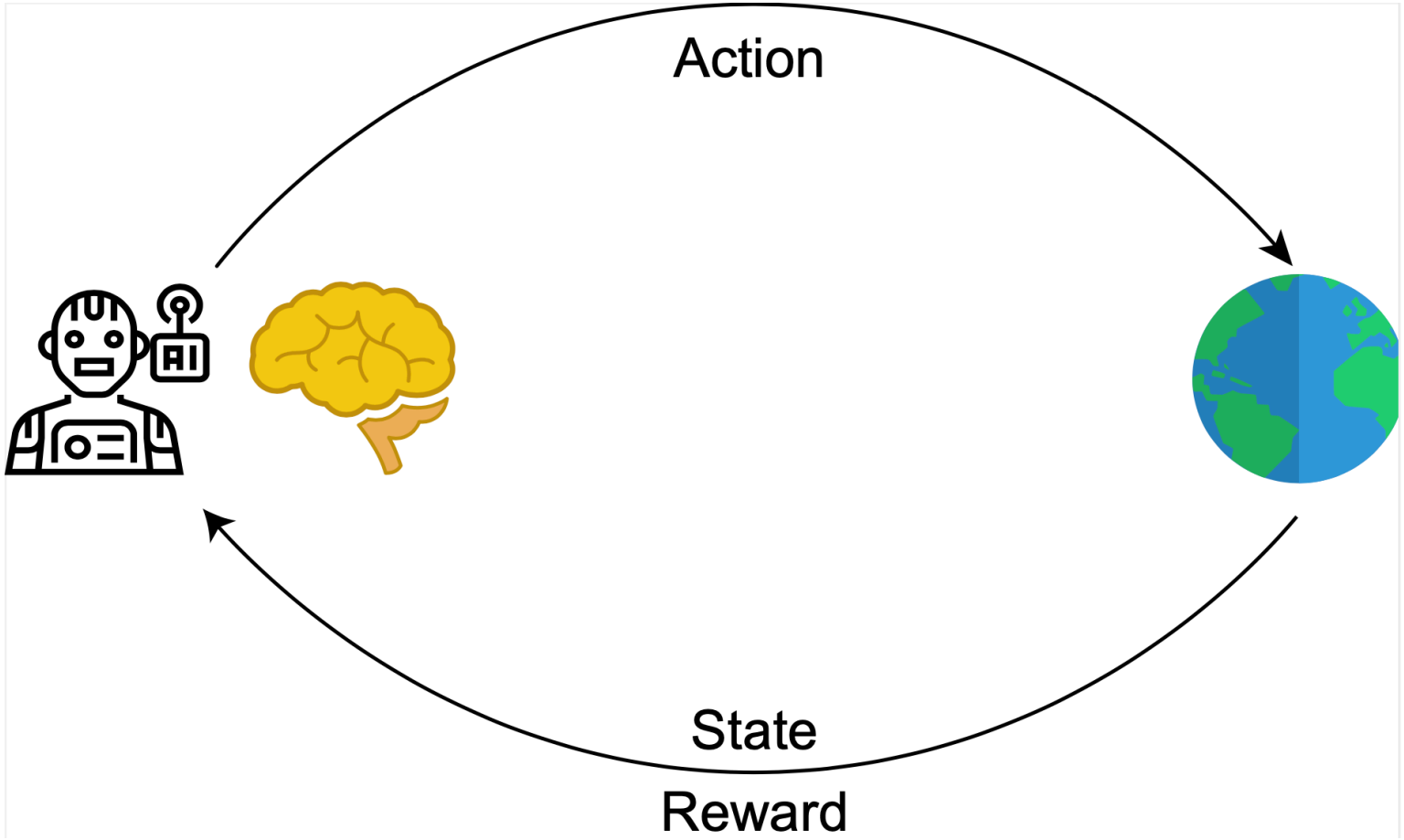
Learning, decision making and complexity

By Harvey Huang

Conversation started with…

- The reinforcement learning community started to talk about (and mathematically model) risk-seeking and risk-averse decision-making.
  - I was puzzled by some concepts.
- E.g., Section 2.4 https://arxiv.org/abs/1806.06923
- “Unfortunately, Section 2.4 provides a very narrow and biased view of the literature, and in my view reflects poor understanding of the issues. ”
- The goal of this presentation is to summarize the conversation with Peter, and to give a (very) brief overview of how learning, decision making and complexity are linked.

(Deep) reinforcement learning



Components breakdown

- State:
  - Information received from the environment
  - (Discrete, hypothetically clustered): e.g., sunshine vs. rainy vs. cloudy
  - (Continuous and highly complex):e.g., 64 x 64 pixels from Atari.
- Action:
  - Eat ice-cream vs. drink hot chocolate.
- Policy:
  - Maps state to actions (e.g., I eat ice-cream when sunshine and drink hot chocolate when rainy)
  - “learning” <=> learn a policy.
- Reward function:
  - (Really) important for policy evaluation and improvement (so that I know I should not eat ice-cream when its rainy and cold).

Optimal action

- In most cases, the policy is of the utility maximization type.
  - Optimal action = argmax(state-action value) or argmax(state-action probability)
- Regardless how complicated the algorithms are, in most cases, RL agents make decisions based on this formula.

Reinforcement learning view on decisions

- “Decisions are guided by ability to associate environmental cues with the outcomes of our chosen actions.”
  - RL research is about “how” this association emerges.
  - Model of the process, not just the outcome.

A vague summary of RL in CS vs. in behavioral studies

	RL in Computer Science	RL in behavioral studies
General goal	Simulation	Fitting
Direction	Forward-looking, (in the classical case) start with an untrained model, training while playing.	Backward-looking, you start with collected behavioral data.
Goal	Train a model to obtain the best performance (e.g., cumulative rewards)	Find and interpret optimally fit model/parameter (e.g., learning rate)
Example convention difference	E-greedy	Soft-max

Decision studies

- In some sense, the max(state-action value) policies are equivalent to the Von Neumann Morgenstern expected utility framework.
  - Utility function ⇔ Value function.
  - Both frameworks attempt to solve an optimization problem in learning and decision-making.
    - What is the optimal policy?
  - (in general) Economists like to take the first order condition whereas the RL community likes to optimize iteratively.
    - Outcome (choice) vs. learning process

RL and decision studies

- Agents choose actions (as if) maximize value/utility function.
  - This is an important assumption.
  - (In my view) not mentioned explicitly enough; often take as granted/convention.
- Humans exhibit “irrational” behavior.
  - Making choices that are inconsistent with what models (max utility/value function) would otherwise predict.

Discipline view on irrational behavior

- RL engineers (in CS) generally don't think this is a big issue
  - why would agents NOT optimize? Unless irrational choices generate better performance (contradicts to "rational").
  - Treat non-optimized choices as
    - "room to improve" and turn into sophisticated algorithms (or complicated neural network structure).
    - or it is just part of the learning process.
- RL behavioral studies acknowledge those non-optimized choices.
  - Treat them as "noises" or "uncertainties", modelled by either learning (prediction error * learning rate) or exploration (echo diffusion models) as part of learning process (the beta in the softmax function).
  - Focus on "learning".
  - [*Transforming values into choice*] https://www.nature.com/articles/s41386-021-01126-y
- Decision studies acknowledge irrational choices as well.
  - Build all sorts of models to rationalize the "irrationals".
  - These models try to explain (fit) a set of agents' choices.
  - Focus on "choice outcome" and rationalization.

Problems

- RL
  - Internal “values” are reflected by choices.
  - Lack of attention on decision making. In the past, most algorithms assume max(value function).
  - It looks like that the community has started to pay attention to the “decision part”.
    - But seems cherry-picking from the literature.
    - E.g., max (f(value function)), f: value -> value*
- Decision studies
  - Internal values are reflected by choices as well.
  - Lack of attention on learning process, or the process of getting to the choices.
    - (There is an equilibrium, as long as we can prove/explain it, we don’t care how equilibrium is obtained).

Decision studies

- ``one of the most useful aspects of decision theory is to answer the question: "If I made a decision, how do I know that it is well though through?" ``
- ⇔ policy improvement in the RL framework (CS).
  - reward function matters.

Another dimension: complexity (dimensionality)

- Modelled state: sunshine vs. rainy
  - Actual state: 10 degree C vs. 20 degree C vs. 35 degree C, humid vs. dry, etc..
- Modelled action: eat ice-cream vs. drink hot chocolate.
  - Actual action: go to woolies (thousands of food choices).
- Learning efficiency and choice efficiency.
  - How are humans able to learn and make choices so fast? How does dimension reduction work in humans?
  - Ability to process a LARGE amount of information and make a choice out of a LARGE number of possible choices in a short period of time.
- RL engineers' approach:
  - State/Action abstraction: dimension reduction/compression.