Reinforcement Learning in Psychology and Neuroscience

with thanks to Elliot Ludvig
University of Warwick
Bidirectional Influences

Psychology

Neuroscience

Reinforcement Learning

Artificial Intelligence

Control Theory
Any information processing system can be understood at multiple “levels”

• The Computational Theory Level
  – *What* is being computed?
  – *Why* are these the right things to compute?

• Representation and Algorithm Level
  – *How* are these things computed?

• Implementation Level
  – How is this implemented physically?
Goals for today’s lecture

• To learn:

  • That psychology recognizes two fundamental learning processes, analogous to our prediction and control.

  • That all the ideas in this course are also important in completely different fields: psychology and neuroscience

  • That the details of the TD(\(\lambda\)) algorithm match key features of biological learning
Psychology has identified two primitive kinds of learning

- Classical Conditioning
- Operant Conditioning (a.k.a. Instrumental learning)

Computational theory:

- Classical = Prediction
  - What is going to happen?
- Operant = Control
  - What to do to maximize reward?
Classical Conditioning
Classical Conditioning as Prediction Learning

- Classical Conditioning is the process of learning to predict the world around you
  - Classical Conditioning concerns (typically) the subset of these predictions to which there is a hard-wired response
Pavlov (1901)

- Russian physiologist
- Interested in how learning happened in the brain
- Conditional and Unconditional Stimuli
Is it really predictions?
Maybe Contiguity?

- Foundational principle of classical associationism (back to Aristotle)
  - Contiguity = Co-occurrence
  - Sufficient for association?
Contiguity Problems

- Unnecessary:
  - Conditioned Taste Aversion
- Insufficient:
  - Blocking
  - Contingency Experiments
Blocking

Phase 1

Light comes to cause salivation

Phase 2

Will sound come to cause salivation? No.

Learning about the sound in Phase 2 does not occur because it is blocked by the association formed in Phase 1.
Rescorla-Wagner Model (1972)

- Computational model of conditioning
  - Widely cited and used
- Learning as violation of expectations
  - As in linear supervised learning (LMS, p2)
  - TD learning is a real-time extension of this same idea
Operant Learning

• The natural learning process directly analogous to reinforcement learning

• Control! What response to make when?
Thorndike’s Puzzle Box (1910)
Law of Effect

- “Of several responses made to the same situation, those which are accompanied by or closely followed by satisfaction to the animal will, other things being equal, be more firmly connected with the situation, so that, when it recurs, they will be more likely to recur...” - Thorndike (1911), p. 244
Operant Chambers
Complex Cognition
Any information processing system can be understood at multiple “levels”

• The Computational Theory Level
  – *What* is being computed?
  – *Why* are these the right things to compute?

• Representation and Algorithm Level
  – *How* are these things computed?

• Implementation Level
  – How is this implemented physically?

David Marr, 1972
Learn to predict discounted sum of upcoming reward through TD with linear function approximation

The TD error is calculated as:

$$\delta_t \doteq R_{t+1} + \gamma \hat{v}(S_{t+1}, \theta) - \hat{v}(S_t, \theta)$$
TD(\(\lambda\)) algorithm/model/neuron

\[ \dot{w}_i \sim \delta \cdot e_i \]

\[ \sum_i w_i \cdot x_i \]

Value of state or action

States or Features

Reward

TD Error

Eligibility Trace

TD Error
Brain reward systems

What signal does this neuron carry?

Honeybee Brain

VUM Neuron

Hammer, Menzel
Dopamine

- Small-molecule Neurotransmitter
  - Diffuse projections from mid-brain throughout the brain

Key Idea: dopamine responding = TD error

Dopamine neurons are a core system that controls motivation in the brain (Fig. 41). All addictive drugs act by increasing the level of dopamine activity, and when enough dopamine neurons die, the symptoms of Parkinson's disease appear, which include motor tremor, difficulty initiating actions and, eventually, anhedonia, the complete loss of pleasure in any activity, ending in catatonia, a complete lack of movement and responsiveness. But when the dopamine cells are behaving normally, they provide brief bursts of dopamine to the cortex and other brain areas when an unexpected reward occurs, and a diminution of activity when...
What does Dopamine Do?

- Hedonic Impact
- Motivation
- Motor Activity
- Attention
- Novelty
- Learning
TD Error = Dopamine

Schultz et al., (1997); Montague et al. (1996)
Dopamine neurons signal the error/change in prediction of reward
No prediction
Reward occurs

Reward predicted
Reward occurs

Reward predicted
No reward occurs

\[ \delta_t = R_{t+1} + \gamma \hat{v}_{t+1} - \hat{v}_t \]
The theory that *Dopamine = TD error* is the *most important interaction ever* between AI and neuroscience.
Goals for today’s lecture

• To learn:

  • That psychology recognizes two fundamental learning processes, analogous to our prediction and control.

  • That all the ideas in this course are also important in completely different fields: psychology and neuroscience

  • That the details of the TD(\(\lambda\)) algorithm match key features of biological learning
What have you learned about in this course (without buzzwords)?

• “Decision-making over time to achieve a long-term goal”
  - includes learning and planning
  - makes plain why value functions are so important
  - makes plain why so many fields care about these algorithms
    • AI
    • Control theory
    • Psychology and Neuroscience
    • Operations Research
    • Economics
  - all involve decision, goals, and time...
    • the essence of...
Bidirectional Influences

Psychology → Reinforcement Learning

Neuroscience → Reinforcement Learning

Reinforcement Learning → Artificial Intelligence

Reinforcement Learning → Control Theory