

CS394R
Reinforcement Learning:
Theory and Practice

Scott Niekum and Peter Stone

Department of Computer Science
The University of Texas at Austin

BE a reinforcement learner

BE a reinforcement learner

- You, as a class, act as a learning agent

BE a reinforcement learner

- You, as a class, act as a learning agent
- **Actions:** Wave, Stand, Clap

BE a reinforcement learner

- You, as a class, act as a learning agent
- **Actions:** Wave, Stand, Clap
- **Observations:** colors, reward

BE a reinforcement learner

- You, as a class, act as a learning agent
- **Actions:** Wave, Stand, Clap
- **Observations:** colors, reward
- **Goal:** Find an optimal *policy*

BE a reinforcement learner

- You, as a class, act as a learning agent
- **Actions:** Wave, Stand, Clap
- **Observations:** colors, reward
- **Goal:** Find an optimal *policy*
 - Way of selecting actions that gets you the most reward

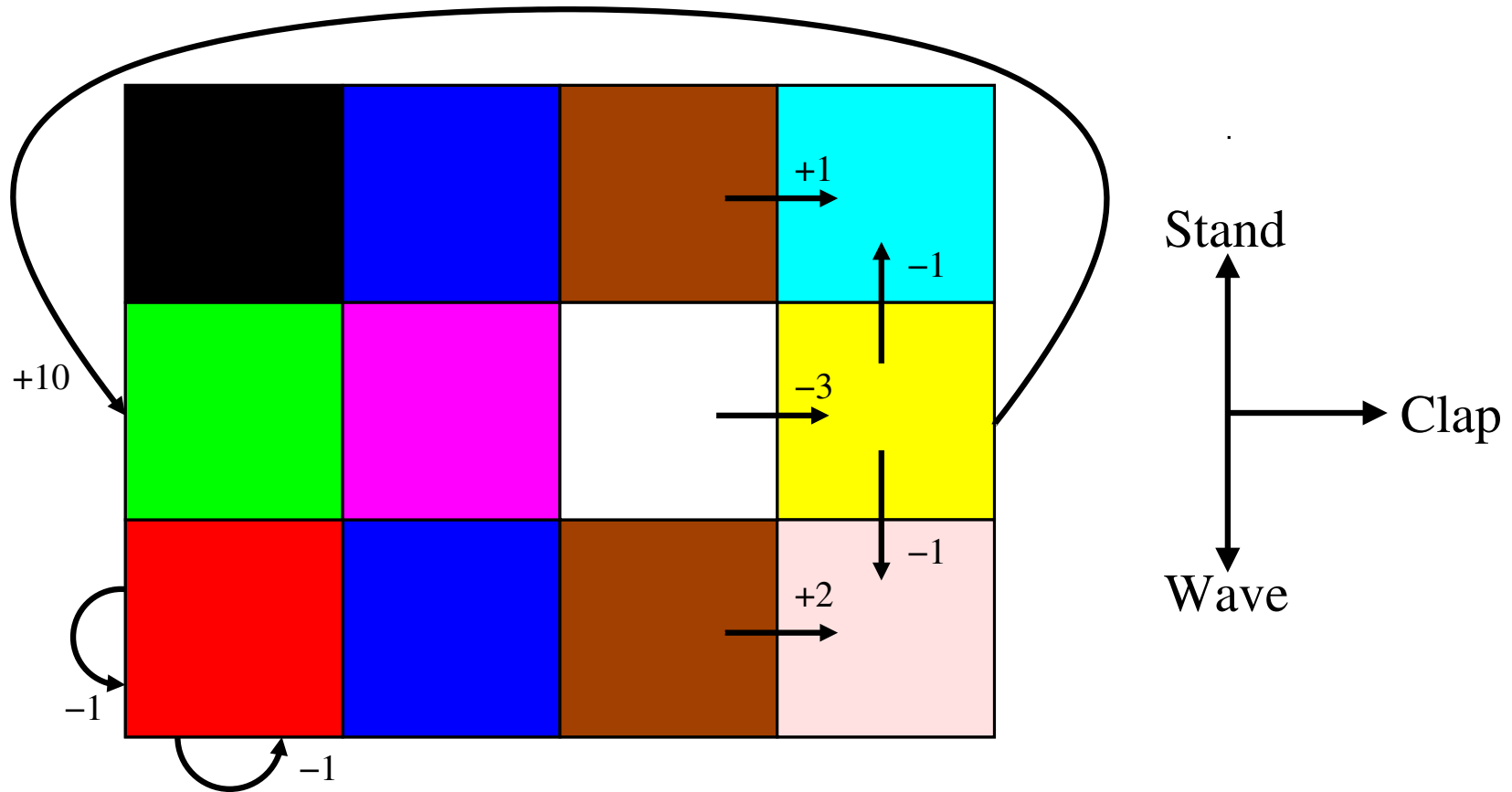
How did you do it?

How did you do it?

- What is your policy?
- What does the world look like?

How did you do it?

- What is your policy?
- What does the world look like?



Formalizing What Just Happened

Knowns:

Formalizing What Just Happened

Knowns:

- $\mathcal{O} = \{\text{Blue, Red, Green, Black, \dots}\}$
- Rewards in \mathbb{R}
- $\mathcal{A} = \{\textit{Wave, Clap, Stand}\}$

Formalizing What Just Happened

Knowns:

- $\mathcal{O} = \{\text{Blue, Red, Green, Black, } \dots\}$
- Rewards in \mathbb{R}
- $\mathcal{A} = \{\textit{Wave, Clap, Stand}\}$

$o_0, a_0, r_0, o_1, a_1, r_1, o_2, \dots$
--

Formalizing What Just Happened

Knowns:

- $\mathcal{O} = \{\text{Blue, Red, Green, Black, \dots}\}$
- Rewards in \mathbb{R}
- $\mathcal{A} = \{\textit{Wave, Clap, Stand}\}$

$o_0, a_0, r_0, o_1, a_1, r_1, o_2, \dots$
--

Unknowns:

Formalizing What Just Happened

Knowns:

- $\mathcal{O} = \{\text{Blue, Red, Green, Black, \dots}\}$
- Rewards in \mathbb{R}
- $\mathcal{A} = \{\text{Wave, Clap, Stand}\}$

$o_0, a_0, r_0, o_1, a_1, r_1, o_2, \dots$
--

Unknowns:

- $\mathcal{S} = 4 \times 3$ grid
- $\mathcal{R} : \mathcal{S} \times \mathcal{A} \mapsto \mathbb{R}$
- $\mathcal{T} = \mathcal{S} \mapsto \mathcal{O}$
- $\mathcal{P} : \mathcal{S} \times \mathcal{A} \mapsto \mathcal{S}$

Formalizing What Just Happened

Knowns:

- $\mathcal{O} = \{\text{Blue, Red, Green, Black, } \dots\}$
- Rewards in \mathbb{R}
- $\mathcal{A} = \{\text{Wave, Clap, Stand}\}$

$o_0, a_0, r_0, o_1, a_1, r_1, o_2, \dots$

Unknowns:

- $\mathcal{S} = 4 \times 3$ grid
- $\mathcal{R} : \mathcal{S} \times \mathcal{A} \mapsto \mathbb{R}$
- $\mathcal{T} = \mathcal{S} \mapsto \mathcal{O}$
- $\mathcal{P} : \mathcal{S} \times \mathcal{A} \mapsto \mathcal{S}$

$s_0, o_0, a_0, r_0, s_1, o_1, a_1, r_1, s_2, o_2, \dots$

Formalizing What Just Happened

Knowns:

- $\mathcal{O} = \{\text{Blue, Red, Green, Black, \dots}\}$
- Rewards in \mathbb{R}
- $\mathcal{A} = \{\text{Wave, Clap, Stand}\}$

$o_0, a_0, r_0, o_1, a_1, r_1, o_2, \dots$

Unknowns:

- $\mathcal{S} = 4 \times 3$ grid
- $\mathcal{R} : \mathcal{S} \times \mathcal{A} \mapsto \mathbb{R}$
- $\mathcal{T} = \mathcal{S} \mapsto \mathcal{O}$
- $\mathcal{P} : \mathcal{S} \times \mathcal{A} \mapsto \mathcal{S}$

$s_0, o_0, a_0, r_0, s_1, o_1, a_1, r_1, s_2, o_2, \dots$

$$o_i = \mathcal{T}(s_i)$$

Formalizing What Just Happened

Knowns:

- $\mathcal{O} = \{\text{Blue, Red, Green, Black, } \dots\}$
- Rewards in \mathbb{R}
- $\mathcal{A} = \{\text{Wave, Clap, Stand}\}$

$o_0, a_0, r_0, o_1, a_1, r_1, o_2, \dots$

Unknowns:

- $\mathcal{S} = 4 \times 3$ grid
- $\mathcal{R} : \mathcal{S} \times \mathcal{A} \mapsto \mathbb{R}$
- $\mathcal{T} = \mathcal{S} \mapsto \mathcal{O}$
- $\mathcal{P} : \mathcal{S} \times \mathcal{A} \mapsto \mathcal{S}$

$s_0, o_0, a_0, r_0, s_1, o_1, a_1, r_1, s_2, o_2, \dots$

$$o_i = \mathcal{T}(s_i)$$

$$r_i = \mathcal{R}(s_i, a_i)$$

Formalizing What Just Happened

Knowns:

- $\mathcal{O} = \{\text{Blue, Red, Green, Black, \dots}\}$
- Rewards in \mathbb{R}
- $\mathcal{A} = \{\text{Wave, Clap, Stand}\}$

$o_0, a_0, r_0, o_1, a_1, r_1, o_2, \dots$

Unknowns:

- $\mathcal{S} = 4 \times 3$ grid
- $\mathcal{R} : \mathcal{S} \times \mathcal{A} \mapsto \mathbb{R}$
- $\mathcal{T} = \mathcal{S} \mapsto \mathcal{O}$
- $\mathcal{P} : \mathcal{S} \times \mathcal{A} \mapsto \mathcal{S}$

$s_0, o_0, a_0, r_0, s_1, o_1, a_1, r_1, s_2, o_2, \dots$

$$o_i = \mathcal{T}(s_i)$$

$$r_i = \mathcal{R}(s_i, a_i)$$

$$s_{i+1} = \mathcal{P}(s_i, a_i)$$

This Course

- Reinforcement Learning theory (start)

This Course

- Reinforcement Learning theory (start)
- Reinforcement Learning in practice (end)

The Big Picture

- AI

The Big Picture

- AI \rightarrow ML

The Big Picture

- AI \longrightarrow ML \longrightarrow RL

The Big Picture

- AI \longrightarrow ML \longrightarrow RL
- Types of Machine Learning

The Big Picture

- AI \longrightarrow ML \longrightarrow RL
- Types of Machine Learning
 - **Supervised learning:** learn from labeled examples

The Big Picture

- AI \longrightarrow ML \longrightarrow RL

- Types of Machine Learning

 - **Supervised learning:** learn from labeled examples

 - **Unsupervised learning:** cluster unlabeled examples

The Big Picture

- AI \longrightarrow ML \longrightarrow RL

- Types of Machine Learning

Supervised learning: learn from labeled examples

Unsupervised learning: cluster unlabeled examples

Reinforcement learning: learn from interaction

The Big Picture

- AI \longrightarrow ML \longrightarrow RL

- Types of Machine Learning

Supervised learning: learn from labeled examples

Unsupervised learning: cluster unlabeled examples

Reinforcement learning: learn from interaction

- Defined by the problem

The Big Picture

- AI \longrightarrow ML \longrightarrow RL

- Types of Machine Learning

Supervised learning: learn from labeled examples

Unsupervised learning: cluster unlabeled examples

Reinforcement learning: learn from interaction

- Defined by the problem

- Many approaches possible (including evolutionary)

The Big Picture

- AI \longrightarrow ML \longrightarrow RL

- Types of Machine Learning

Supervised learning: learn from labeled examples

Unsupervised learning: cluster unlabeled examples

Reinforcement learning: learn from interaction

- Defined by the problem
- Many approaches possible (including evolutionary)
- Book focusses on a particular class of approaches

Reduced Formalism

Knowns:

- $\mathcal{S} = \{\text{Blue, Red, Green, Black, } \dots\}$
- Rewards in \mathbb{R}
- $\mathcal{A} = \{\textit{Wave, Clap, Stand}\}$

$s_0, a_0, r_0, s_1, a_1, r_1, s_2, \dots$

Reduced Formalism

Knowns:

- $\mathcal{S} = \{\text{Blue, Red, Green, Black, } \dots\}$
- Rewards in \mathbb{R}
- $\mathcal{A} = \{\textit{Wave, Clap, Stand}\}$

$s_0, a_0, r_0, s_1, a_1, r_1, s_2, \dots$

Unknowns:

Reduced Formalism

Knowns:

- $\mathcal{S} = \{\text{Blue, Red, Green, Black, } \dots\}$
- Rewards in \mathbb{R}
- $\mathcal{A} = \{\textit{Wave, Clap, Stand}\}$

$s_0, a_0, r_0, s_1, a_1, r_1, s_2, \dots$

Unknowns:

- $\mathcal{R} : \mathcal{S} \times \mathcal{A} \mapsto \mathbb{R}$
- $\mathcal{P} : \mathcal{S} \times \mathcal{A} \mapsto \mathcal{S}$

Reduced Formalism

Knowns:

- $\mathcal{S} = \{\text{Blue, Red, Green, Black, \dots}\}$
- Rewards in \mathbb{R}
- $\mathcal{A} = \{\text{Wave, Clap, Stand}\}$

$$s_0, a_0, r_0, s_1, a_1, r_1, s_2, \dots$$

Unknowns:

- $\mathcal{R} : \mathcal{S} \times \mathcal{A} \mapsto \mathbb{R}$
- $\mathcal{P} : \mathcal{S} \times \mathcal{A} \mapsto \mathcal{S}$

$$r_i = \mathcal{R}(s_i, a_i)$$

$$s_{i+1} = \mathcal{P}(s_i, a_i)$$

This course

- Agent's perspective: only **policy** under control
 - State representation, reward function given
 - Focus on policy algorithms, theoretical analyses

This course

- Agent's perspective: only **policy** under control
 - State representation, reward function given
 - Focus on policy algorithms, theoretical analyses
 - Appeal: program by just specifying goals

This course

- Agent's perspective: only **policy** under control
 - State representation, reward function given
 - Focus on policy algorithms, theoretical analyses
 - Appeal: program by just specifying goals
 - Practice: need to pick the representation, reward

This course

- Agent's perspective: only **policy** under control
 - State representation, reward function given
 - Focus on policy algorithms, theoretical analyses
 - Appeal: program by just specifying goals
 - Practice: need to pick the representation, reward
 - videos

This course

- Agent's perspective: only **policy** under control
 - State representation, reward function given
 - Focus on policy algorithms, theoretical analyses
 - Appeal: program by just specifying goals
 - Practice: need to pick the representation, reward
 - videos
- Methodical approach
 - Solid foundation rather than comprehensive coverage

This course

- Agent's perspective: only **policy** under control
 - State representation, reward function given
 - Focus on policy algorithms, theoretical analyses
 - Appeal: program by just specifying goals
 - Practice: need to pick the representation, reward
 - videos
- Methodical approach
 - Solid foundation rather than comprehensive coverage
 - RL reading group

Syllabus

- Available on-line

Assignments

- Join edX!

Assignments

- Join edX!
- Read Chapters 2 and 3 (and 1 if you haven't)

Assignments

- Join edX!
- Read Chapters 2 and 3 (and 1 if you haven't)
- Send a reading response by 5pm Monday

Assignments

- Join edX!
- Read Chapters 2 and 3 (and 1 if you haven't)
- Send a reading response by 5pm Monday
- Start on your first programming assignment

Assignments

- Join edX!
- Read Chapters 2 and 3 (and 1 if you haven't)
- Send a reading response by 5pm Monday
- Start on your first programming assignment