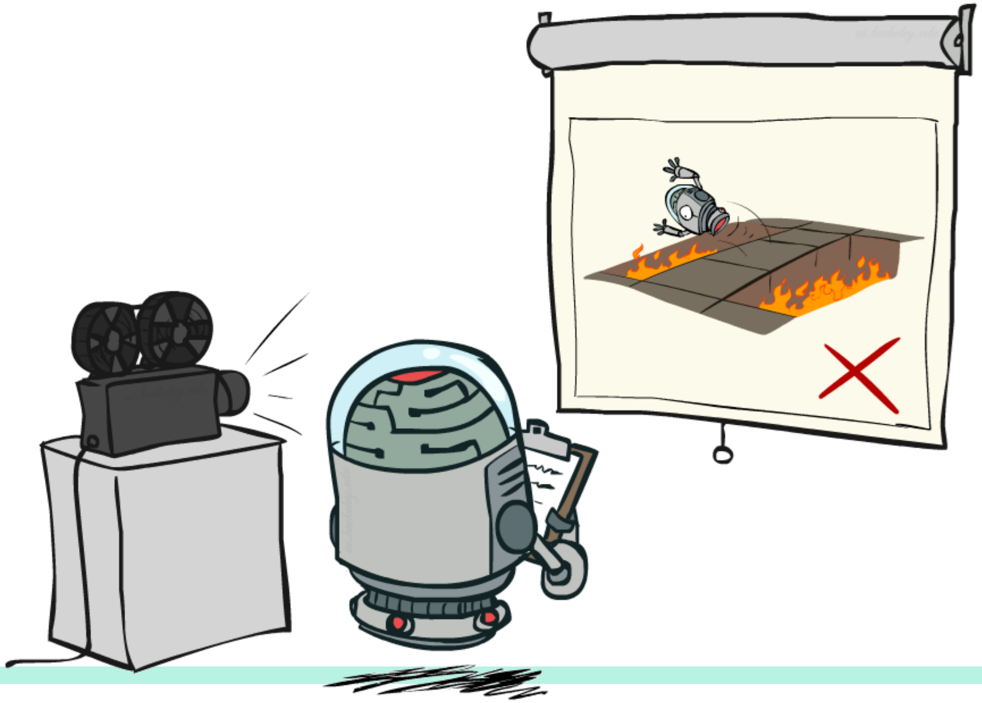
Advanced Topics in AI

Direct Evaluation



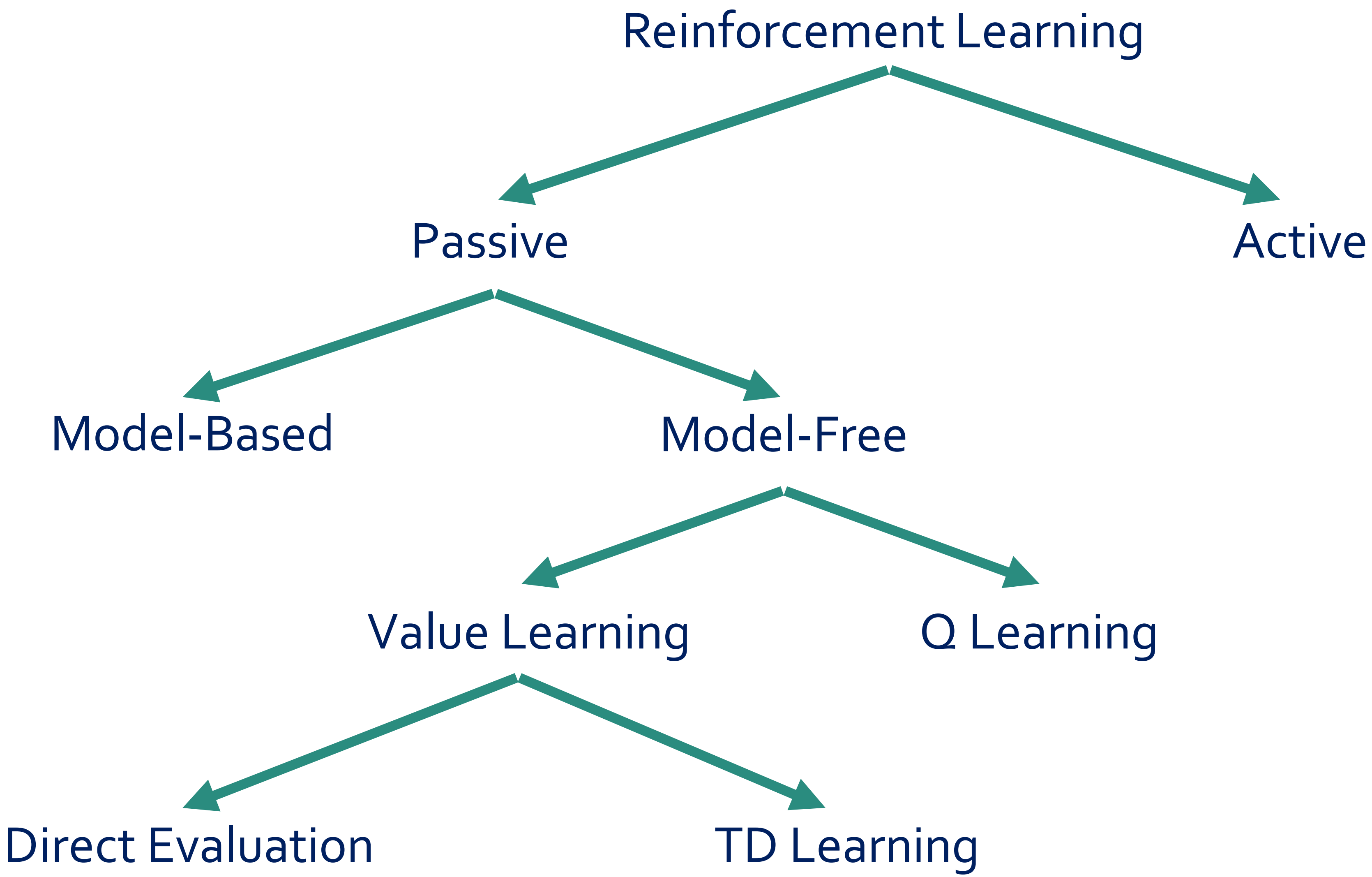
Instructor: Prof. Dr. techn. Wolfgang Nejdl

Leibniz University Hannover

[These slides were created by Dan Klein and Pieter Abbeel for CS188 Intro to AI at UC Berkeley. All materials are available at http://ai.berkeley.edu.]

Co-financed by the Connecting Europe Facility of the European Union

Reinforcement Learning Taxonomy



Reinforcement Learning

Passive

Active

Model-Based

Model-Free

Value Learning

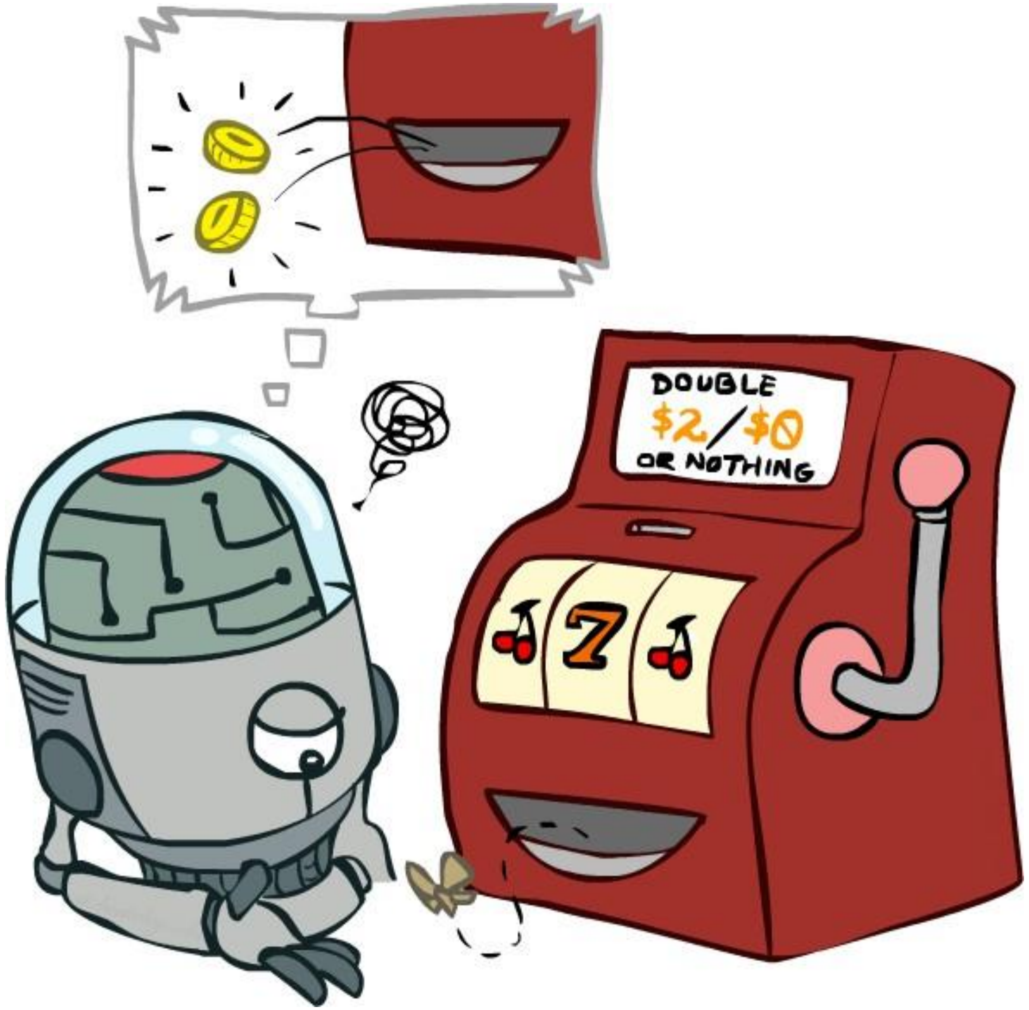
Q Learning

Direct Evaluation

TD Learning

Co-financed by the Connecting Europe Facility of the European Union

Model-Free Learning



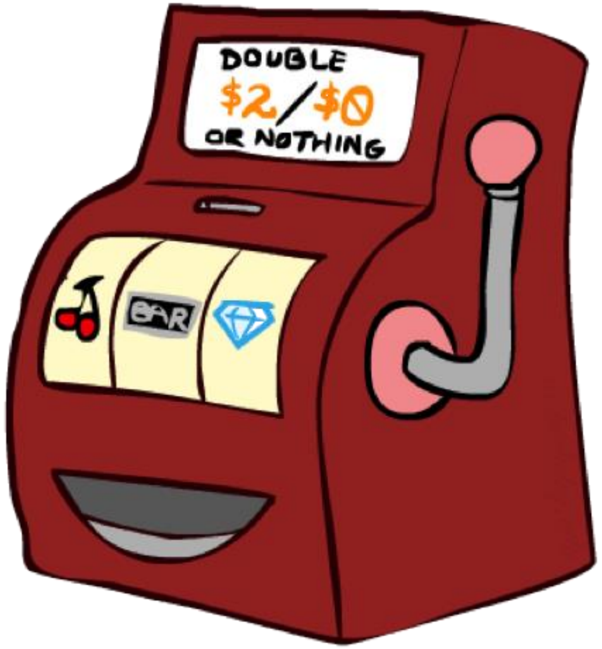
Co-financed by the Connecting Europe Facility of the European Union

Direct Evaluation

- Goal: Compute values for each state under π

- Idea: Average together observed sample values
  - Act according to π
  - Every time you visit a state, write down what the sum of discounted rewards turned out to be
  - Average those samples

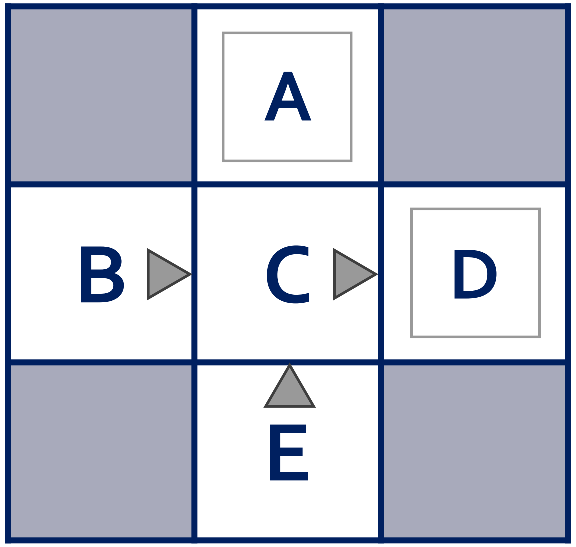
- This is called direct evaluation



Co-financed by the Connecting Europe Facility of the European Union

Example: Direct Evaluation

Input Policy π



Assume: $\gamma = 1$

Observed Episodes (Training)

Episode 1

B, east, C, -1
C, east, D, -1
D, exit, x, +10

Episode 2

B, east, C, -1
C, east, D, -1
D, exit, x, +10

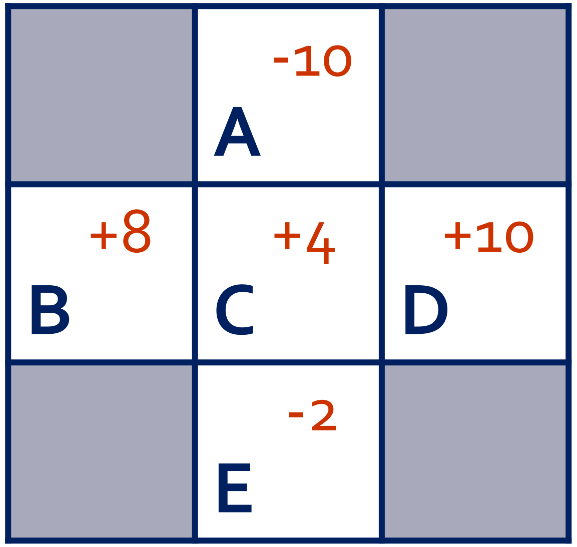
Episode 3

E, north, C, -1
C, east, D, -1
D, exit, x, +10

Episode 4

E, north, C, -1
C, east, A, -1
A, exit, x, -10

Output Values



$$\text{sample}_i(s) = \sum_t \gamma^t R^t$$

$$V(s) \approx \frac{1}{N} \sum_i \text{sample}_i(s)$$

Co-financed by the Connecting Europe Facility of the European Union

Quiz: Direct Evaluation

Observed (s, a, s', R) Transitions

Episode 1

E, north, C, -1
C, east, D, -1
D, exit, x, +10

Episode 2

C, east, A, -1
A, exit, x, -5

Episode 3

B, east, C, -1
C, east, D, -1
D, exit, x, +10

$$\text{sample}_i(s) = \sum_t \gamma^t R^t$$

$$V(s) \approx \frac{1}{N} \sum_i \text{sample}_i(s)$$

Assume: $\gamma = 1$

What is value of state C via Direct Evaluation?

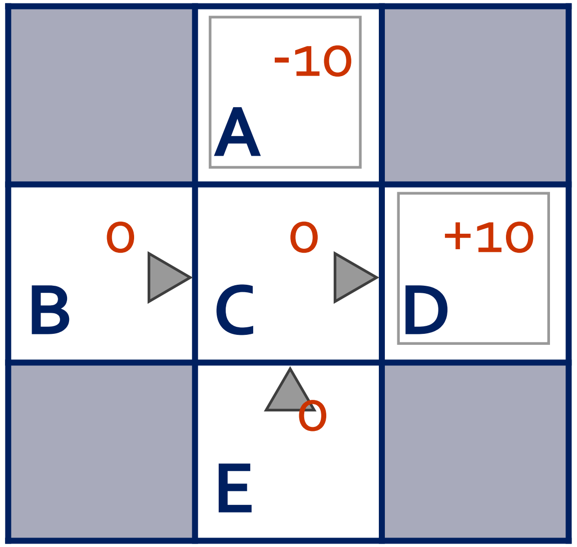
Co-financed by the Connecting Europe Facility of the European Union

Problems with Direct Evaluation

- What's good about direct evaluation?
  - It's easy to understand
  - It doesn't require any knowledge of T, R
  - It eventually computes the correct average values, using just sample transitions

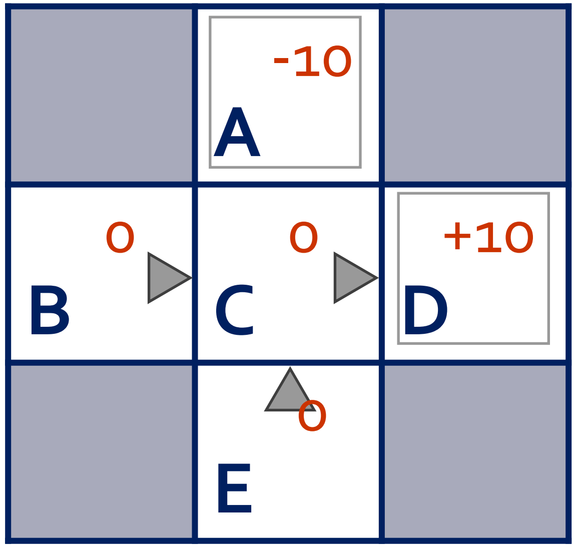
- What bad about it?
  - It wastes information about state connections
  - Need to have all episodes ahead of time (cannot "stream" in transitions)

Output Values



Co-financed by the Connecting Europe Facility of the European Union

Problems with Direct Evaluation



Observed Transitions (s, a, s', R)

Episode 1

E (home),	study,	C (know material),	0
C (know material),	go to exam,	D (pass exam),	0
D (pass exam),	exit,	x,	+10

Episode 2

B (library),	study,	C (know material),	0
C (know material),	go to exam,	A (miss bus & fail exam),	0
A (fail exam),	exit,	x,	-10

Is studying in the library a bad idea?

Co-financed by the Connecting Europe Facility of the European Union

Direct Evaluation

- Goal: Compute values for each state under π
- Idea: Average together observed sample values
  - Act according to π
  - Every time you visit a state, write down what the sum of discounted rewards turned out to be:

$$\text{sample}_i(s) = \sum_t \gamma^t R^t$$

  - Average those samples:

$$V(s) \approx \frac{1}{N} \sum_i \text{sample}_i(s)$$

- This is called direct evaluation

Co-financed by the Connecting Europe Facility of the European Union

Exponential Moving Average

- Traditional Average: $AVG(x) = \frac{1}{N}\sum_n x_n$
  - Need to have all N samples at once (cannot “stream” in samples)

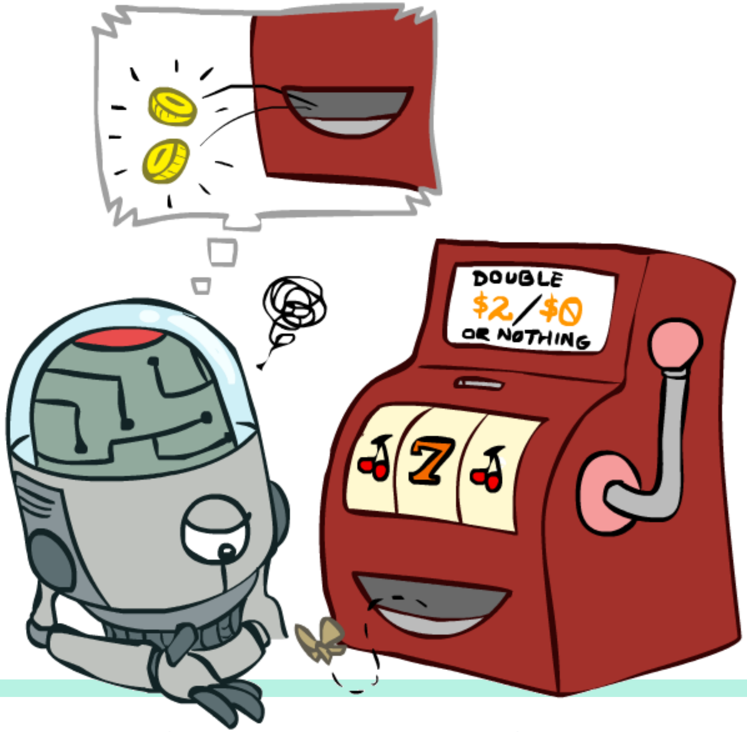
- Exponential moving average
  - The running interpolation update: $\bar{x}_n = (1-\alpha)\cdot\bar{x}_{n-1} + \alpha\cdot x_n$
  - Makes recent samples more important: $\bar{x}_n = \frac{x_n+(1-\alpha)\cdot x_{n-1}+(1-\alpha)^2\cdot x_{n-2}+\ \ldots}{1+(1-\alpha)+(1-\alpha)^2+\ \ldots}$
  - Forgets about the past samples (how quickly depends on α)

- Decreasing learning rate (alpha) can give converging averages

Co-financed by the Connecting Europe Facility of the European Union

Advanced Topics in AI

Next: Temporal Difference Value Learning



Instructor: Prof. Dr. techn. Wolfgang Nejdl

Leibniz University Hannover

[These slides were created by Dan Klein and Pieter Abbeel for CS188 Intro to AI at UC Berkeley. All materials are available at http://ai.berkeley.edu.]

Co-financed by the Connecting Europe Facility of the European Union