INTRO TO RL

Book Chapter: 1

Reinforcement Learning: An Introduction Richard S. Sutton and Andrew G. Barto

http://incompleteideas.net/book/the-book-2nd.html

RL – Origins…

RL-Example: 2 agents

- Gazelle struggles to its feet minutes after being born. Half an hour later it is running fast!
- Robot: a) enter a new room in search of trash or b) start trying to find its way to its battery recharging station. Decision based on: the current charge level and how quickly and easily it has been able to find the recharger in the past.

What do they have In common???????

RL-Example: agent + experience to improve performance over time

Agent:

1. seeks to achieve a goal *despite* uncertainty about its environment.

2. its actions may affect the future state of the environment (e.g., the robot's next location and the future charge level of its battery)

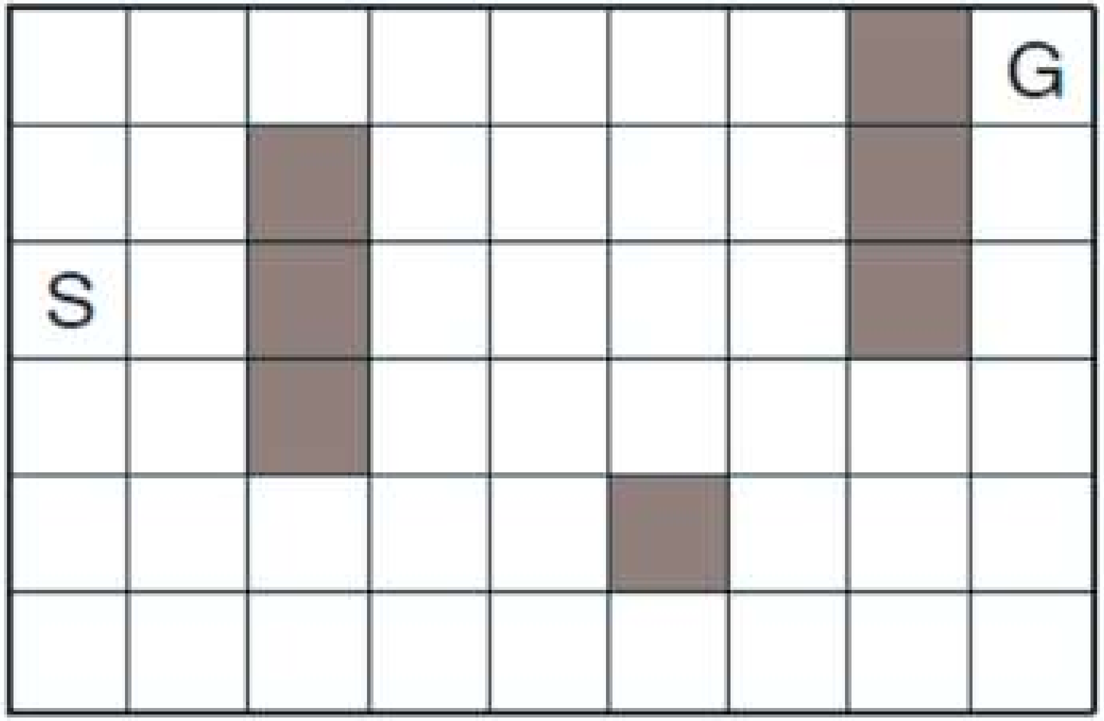
…MEANING… affecting next actions and opportunities available to the agent.

Correct choice: taking into account indirect, delayed consequences of actions (may require foresight or planning).

2 examples: effects of actions cannot be fully predicted: the agent must monitor its environment frequently and react appropriately.

RL simultaneously means:

1. A problem;
2. A class of solution methods that work well on the problem;
3. The field that studies this problem and its solution methods.



							▨	G
		▨					▨	
S		▨					▨	
		▨						
					▨			

RL - **Learning agent**

Learning agent must be able to:

1. Sensation -Sense the state of its environment to some extent;
2. Action-Must be able to take actions that affect the state;
3. Goal-Must have a explicit goal or goals relating to the state of the environment.

State, informal definition

whatever information is available to the agent about its environment.

RL: learning from interaction

- RL: Much more focused on goal-directed learning from interaction than other approaches to ML.

RL: learning what to do (how to map situations to actions) so as to maximize a numerical reward signal.

Learner

- Not told which actions to take;
- Must discover which actions ⇨ reward by trying them;
- Actions may affect the immediate reward but also the next situation (-> all subsequent rewards).

RL: learning from interaction

2 important distinguishing features of RL:

1. Trial-and-error search;
2. Delayed reward.

Is RL =Unsupervised learning???????

3 possible ML paradigms: RL and...

1.Supervised learning: learning from a training set of labeled examples provided by an external supervisor. Object of this kind of learning: for the system to extrapolate, or generalize, its responses so that it acts correctly in situations not present in the training set.

ALONE IS NOT ADEQUATE FOR LEARNING FROM INTERACTION –WHY????

3 possible ML paradigms: RL and…

ALONE IS NOT ADEQUATE FOR LEARNING FROM INTERACTION –WHY????

- In interactive problems it is often impractical to obtain examples of desired behavior that are both correct and representative of all the situations;
- In uncharted territory: an agent must be able to learn from its own experience.

3 possible ML paradigms: RL and...

2.Unsupervised learning: typically about finding structure hidden in collections of unlabeled data.

Yes, RL does not rely on examples of correct behavior

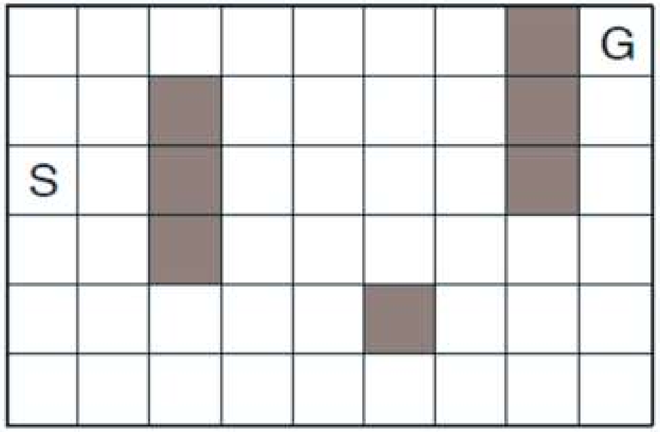
BUT RL is trying to maximize a reward signal instead of trying to find hidden structure.

Unsupervised learning by itself does not address the RL problem of maximizing a reward signal.

RL Challenge: trade-off exploitation vs. Exploration

- Agent must prefer actions found to be effective: reward; ->exploit
- To discover such actions, the agent has to try new actions->explore

Dilemma: **neither can be pursued exclusively without failing at the task**



RL Challenge: trade-off exploitation vs. Exploration

- Agent: try a variety of actions and progressively favor those that appear to be the best;
- Trying just once? On a stochastic task, each action must be tried many times to gain a reliable estimate of its expected reward;
- The agent has to operate despite significant uncertainty about the environment it faces.

Key feature of RL: it explicitly considers the whole problem of a goal-directed agent interacting with an uncertain environment.

Agent + interaction+ environment

4 Elements of RL: 1st

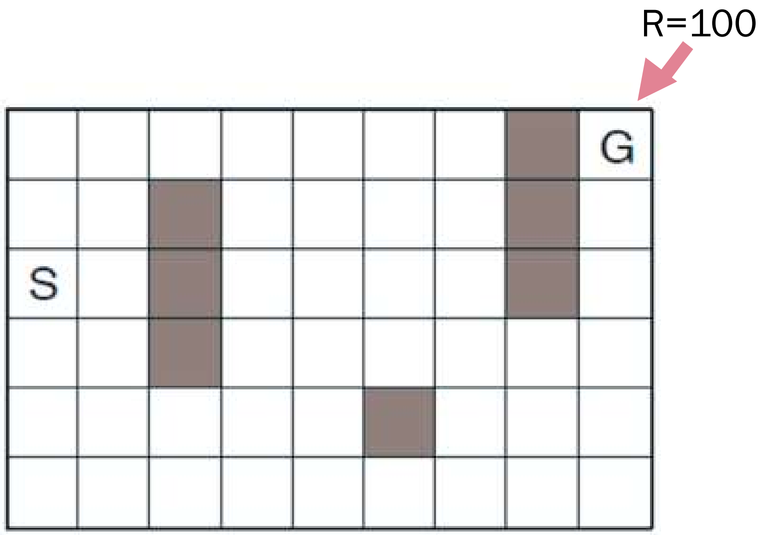
1.Policy: the learning agent’s way of behaving at a given time. A mapping from perceived states of the environment to actions to be taken when in those states (in psychology: a set of stimulus–response rules or associations).

In general, policies may be stochastic, specifying probabilities for each action.

Policy may be a simple function or lookup table, or it may involve extensive computation such as a search process.

Core of a RL: it alone is sufficient to determine behavior.

What is a good policy here?



4 Elements of RL: 2nd

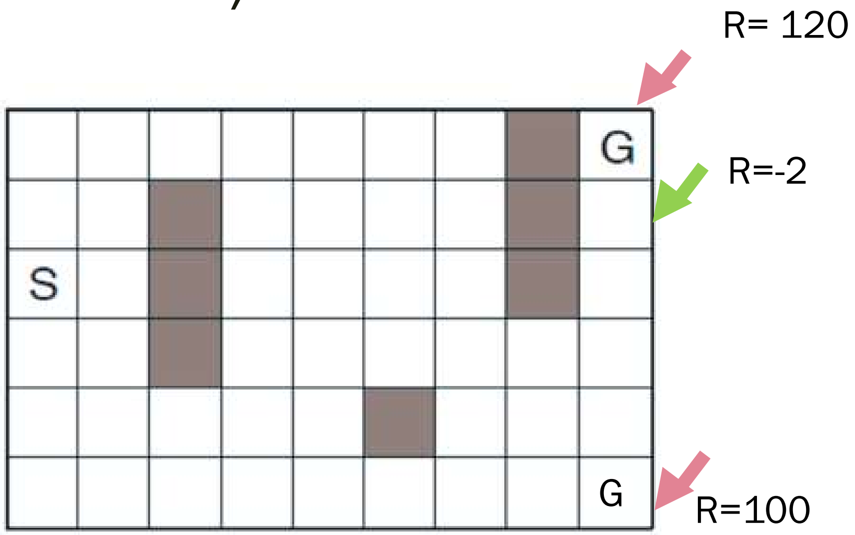
2. Reward signal **Intuition: Pleasure or Pain**.

Defines the goal of a RL problem. On each time step: environment sends to the RL agent a single number called the reward.

The agent’s sole objective is to maximize the total reward it receives over the long run. The reward signal defines what are the good and bad events for the agent.

In general, reward signals may be stochastic functions of the state of the environment and the actions taken.

What is the best policy here?
Each transition, R=-1



4 Elements of RL: 3rd

3. Value function

Reward signal indicates what is good in an immediate sense.

Value function specifies what is good in the long run. What is the value of a state? the total amount of reward an agent can expect to accumulate over the future, starting from that state. Indicates the long-term desirability of states after taking into account the states that are likely to follow and the rewards available in those states.

4 Elements of RL: 3rd

3. Value function

Indicates the long-term desirability of states after taking into account the states that are likely to follow and the rewards available in those states.

Ex: a state might always yield a low immediate reward but still have a high value because it is regularly followed by other states that yield high rewards. Or the reverse could be true.

Without rewards: no values; Purpose of estimating values: to achieve more reward.

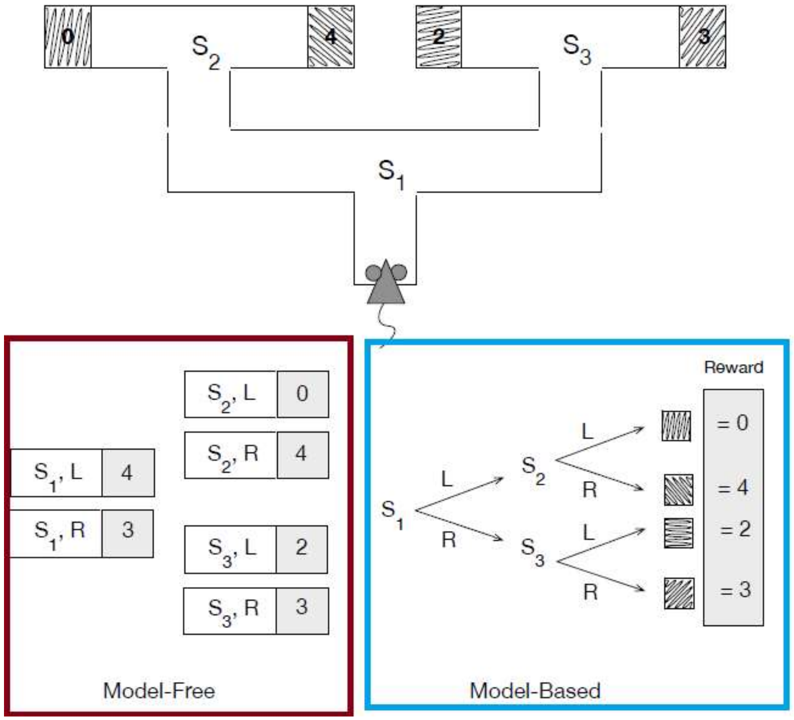
Making and evaluating decisions: we are most concerned with values!

We seek actions that bring about states of highest value, not highest reward: to choose actions that obtain the greatest amount of reward for us over the long run.

Of course: it is harder to determine values than it is to determine rewards;

Rewards are basically given by the environment, but values must be estimated and re-estimated from the sequences of observations an agent makes over its entire lifetime.

Elements of RL: 4th



4. Model of the environment

Something that allows inferences to be made

about how the environment will behave.

Given a state and action, the model might predict

the resultant next state and next reward.

Model-based methods: methods for solving RL problems

that use models and planning;

Model-free methods: explicitly trial-and-error learners

(viewed as almost the opposite of planning).

Fig. RL book, p.365. Model-free strategy relies on stored action values for all the **state–action pairs** obtained over many learning trials. To make decisions, the rat just has to select at each state the action with the **largest action value** for that state.

Model-based strategy, the rat learns an environment model: **state–action-next-state transitions and a reward model** consisting of knowledge of the reward associated with each distinctive goal box. The rat can decide which way to turn at each state by using the model to simulate sequences of action choices to find a path yielding the highest return.