**1.1**

**Reinforcement Learning**

Reinforcement learning is learning what to do—how to map situations to actions—so

as to maximize a numerical reward signal.

All involve interaction between an active decision-making agent and its environment, within

which the agent seeks to achieve a goal despite uncertainty about its environment.

At the same time, in all of these examples the e↵ects of actions cannot be fully predicted;

thus the agent must monitor its environment frequently and react appropriately.

Reinforcement learning is a computational approach to understanding and automating

goal-directed learning and decision making. It is distinguished from other computational

approaches by its emphasis on learning by an agent from direct interaction with its

environment, without requiring exemplary supervision or complete models of the environment.

In our opinion, reinforcement learning is the first field to seriously address the

computational issues that arise when learning from interaction with an environment in

order to achieve long-term goals.

Our focus is on reinforcement learning methods that learn while interacting with the

environment, which evolutionary methods do not do. Methods able to take advantage

of the details of individual behavioral interactions can be much more eficient than

evolutionary methods in many cases.

**A learning agent**

----------------

must be able to sense the state of its environment to some extent and must be able to

take actions that afect the state. facing a learning

agent interacting over time with its environment to achieve a goal. A learning agent

must be able to sense the state of its environment to some extent and must be able to

take actions that a↵ect the state.

All reinforcement learning agents have explicit goals, can sense

aspects of their environments, and can choose actions to influence their environments.

In all of these examples the agent can use its experience to improve its performance

over time

**Markov decision processes**

--------------------------

are intended to include just

these three aspects—sensation, action, and goal—in their simplest possible forms without

trivializing any of them.

We formalize the problem of reinforcement learning using ideas from dynamical systems

theory, specifically, as the optimal control of incompletely-known Markov decision

processes.

**Reinforcement learning is diferent from supervised learning,**

**------------------------------------------------------------**

the kind of learning studied

in most current research in the field of machine learning. Supervised learning is learning

from a training set of labeled examples provided by a knowledgable external supervisor.

Each example is a description of a situation together with a specification—the label—of

the correct action the system should take to that situation, which is often to identify a

category to which the situation belongs. The object of this kind of learning is for the

system to extrapolate, or generalize, its responses so that it acts correctly in situations

not present in the training set. This is an important kind of learning, but alone it is not

adequate for learning from interaction. In interactive problems it is often impractical to

obtain examples of desired behavior that are both correct and representative of all the

situations in which the agent has to act.

**Reinforcement learning is also di↵erent from unsupervised learning,**

---------------------------------------------------------------------

which is typically about finding structure hidden in collections of

unlabeled data.

Although one might be tempted to think of reinforcement learning as a kind of unsupervised learning

because it does not rely on examples of correct behavior, reinforcement learning is trying

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

**1.3 Elements of Reinforcement Learning**

=======================================

Reinforcement learning relies heavily on the concept of state—as input to the policy and

value function, and as both input to and output from the model. Informally, we can

think of the state as a signal conveying to the agent some sense of “how the environment

is” at a particular time..

Beyond the agent and the environment, one can identify four main subelements of a

reinforcement learning system:

a policy, a reward signal, a value function, and, a model of the environment.

**A policy**

--------

defines the learning agent’s way of behaving at a given time. Roughly speaking,

a policy is a mapping from perceived states of the environment to actions to be taken

when in those states. g from perceived states of the environment to actions to be taken

when in those states. It corresponds to what in psychology would be called a set of

stimulus–response rules or associations.

**A reward signal**

---------------

Defines the goal of a reinforcement learning problem. On each time

step, the environment sends to the reinforcement learning 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 thus defines what are the good and bad events for the

agent. In a biological system, we might think of rewards as analogous to the experiences

of pleasure or pain

The reward signal is the primary basis for altering the policy; if an action

selected by the policy is followed by low reward, then the policy may be changed to

select some other action in that situation in the future. In general, reward signals may

be stochastic functions of the state of the environment and the actions taken.

Rewards are basically

given directly by the environment, but values must be estimated and re-estimated from

the sequences of observations an agent makes over its entire lifetime.

**Value Function**

--------------

Whereas the reward signal indicates what is good in an immediate sense, a value

function specifies what is good in the long run. Roughly speaking, the value of a state is

the total amount of reward an agent can expect to accumulate over the future, starting

from that state.

state is whatever information is available to the agent about its environment.

**model of the environment.**

------------------------

This is something that mimics the behavior of the environment, or

more generally, that allows inferences to be made about how the environment will behave.

For example, given a state and action, the model might predict the resultant next state

and next reward. Models are used for planning, by which we mean any way of deciding

on a course of action by considering possible future situations before they are actually

experienced.

Methods for solving reinforcement learning problems that use models and

planning are called model-based methods, as opposed to simpler model-free methods

that are explicitly trial-and-error learners

Anything that the agent cannot change arbitrarily is considered to be part of the environment.

**State**

------

Reinforcement learning relies heavily on the concept of state—as input to the policy and

value function, and as both input to and output from the model. Informally, we can

think of the state as a signal conveying to the agent some sense of “how the environment

is” at a particular time.

The formal definition of state as we use it here is given by

the framework of Markov decision processes presented in Chapter 3. More generally,

however, we encourage the reader to follow the informal meaning and think of the state

as whatever information is available to the agent about its environment. In e↵ect, we

assume that the state signal is produced by some preprocessing system that is nominally

part of the agent’s environment.

**Chapter 3 - Finite Markov Decision Processes**

---------------------------------------------

MDPs are a mathematically idealized form of the reinforcement learning problem

for which precise theoretical statements can be made. We introduce key elements of

the problem’s mathematical structure, such as returns, value functions, and Bellman

equations. We try to convey the wide range of applications that can be formulated as

finite MDPs.

MDPs are meant to be a straightforward framing of the problem of learning from

interaction to achieve a goal.

The learner and decision maker is called the agent.

The thing it interacts with, comprising everything outside the agent, is called the environment.

These interact continually, the agent selecting actions and the environment responding to

these actions and presenting new situations to the agent.1 The environment also gives

rise to rewards, special numerical values that the agent seeks to maximize over time

through its choice of actions.

https://towardsdatascience.com/introduction-to-reinforcement-learning-markov-decision-process-44c533ebf8da

Diagrama

Descripción generada automáticamente

Pour l’etat M: Pour gamma donne.

Carta

Descripción generada automáticamente

Diagrama

Descripción generada automáticamente

R : Reward

G : REndement

Texto, Carta

Descripción generada automáticamente

<https://towardsdatascience.com/reinforcement-learning-markov-decision-process-part-2-96837c936ec3>

Imagen que contiene Carta

Descripción generada automáticamente

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

Interfaz de usuario gráfica, Texto, Sitio web

Descripción generada automáticamente

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

Interfaz de usuario gráfica, Texto, Aplicación, Chat o mensaje de texto

Descripción generada automáticamente

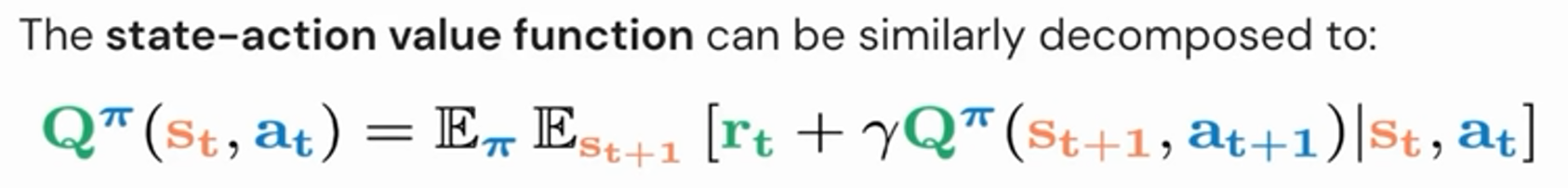
Pour each probability of action(a) in the state (s)

and their value of state associe (Q)

Interfaz de usuario gráfica, Texto

Descripción generada automáticamente

First part is immediate reward.



Interfaz de usuario gráfica, Texto, Aplicación, Word

Descripción generada automáticamente

Knowing optimal Q I know the optimal policy.

Ejemplo 2:

Matrix de probalilite

Diagrama

Descripción generada automáticamente

Ejemplo 3

Calculi de rendement

Diagrama

Descripción generada automáticamente

Diagrama

Descripción generada automáticamente

Texto, Carta

Descripción generada automáticamente

Ejemplo

Calculo Masimo rendement

Texto, Pizarra

Descripción generada automáticamente

Ejemplo

Diagrama

Descripción generada automáticamente

Texto, Carta

Descripción generada automáticamente

Calendario

Descripción generada automáticamente

Diagrama

Descripción generada automáticamente

<https://books.google.ca/books?id=148bEAAAQBAJ&pg=PA334&lpg=PA334&dq=GridworldEnv()&source=bl&ots=ArxjNnzrxF&sig=ACfU3U0e36hOuaYXb5vGh0-dmtQTNygPhQ&hl=es-419&sa=X&ved=2ahUKEwjrq5u_3YX2AhXPlYkEHYSeBksQ6AF6BAgSEAM#v=onepage&q=GridworldEnv()&f=false>

<https://deeplearning.neuromatch.io/tutorials/W3D2_BasicReinforcementLearning/student/W3D2_Tutorial1.html>