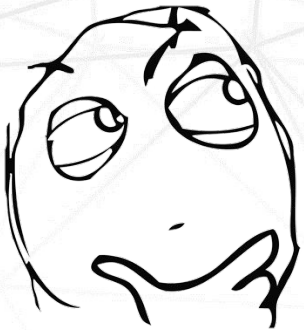
Introducció a l'aprenentatge per reforç

Assignatura: Sistemes d'aprenentatge automàtic

Professor: Ramon Mateo Navarro

vedrunapro
Vall Terrassa

Qué és l'aprenentatge per reforç?



vedrunapro
Vall Terrassa

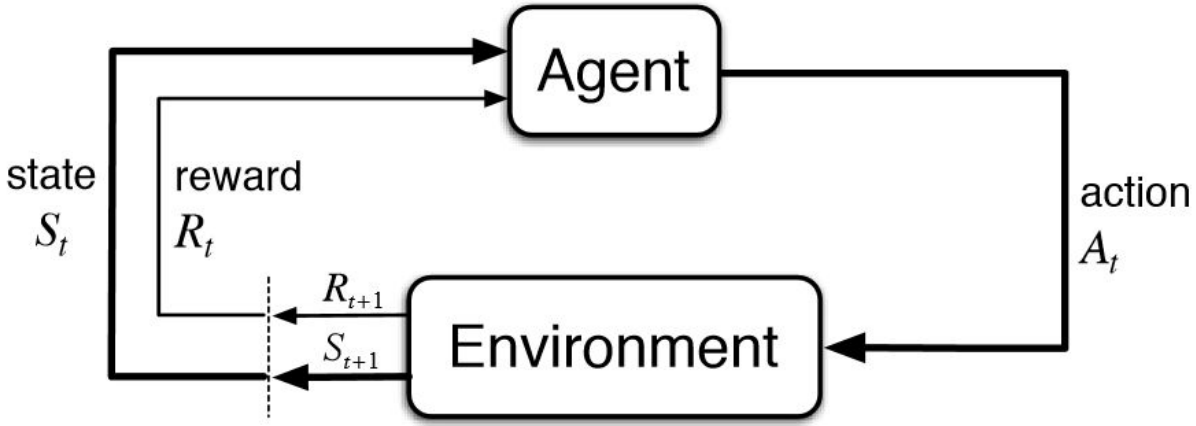
Concepte

L'aprenentatge per reforç és un tipus d'aprenentatge automàtic en què un agent aprèn a prendre decisions sequencialment per aconseguir un objectiu específic a través de la interacció amb un entorn. Aquest tipus d'aprenentatge està inspirat en la psicologia conductual, especialment en la idea de reforçament: l'agent aprèn a partir de les conseqüències de les seves accions, ja siguin recompenses o penalitzacions, amb l'objectiu de maximitzar alguna forma de "recompensa acumulada" al llarg del temps.

Concepte

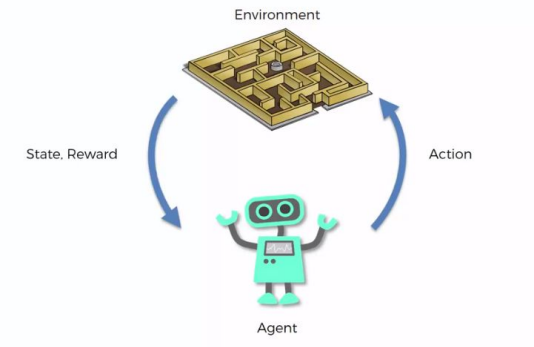

Aprendizaje
Supervisado
Aprendizaje no
Supervisado
Aprendizaje
por Refuerzo

Concepte



Que és un agent?

En l'aprenentatge per reforç, l'agent és l'entitat que pren les decisions i interacciona amb l'entorn per aconseguir un objectiu específic. L'agent pot ser un programa d'ordinador, un robot físic, un sistema autònom, o fins i tot un organisme biològic.



Exemples de casos d'ús

Agents de Jocs d'Ordinador:

- Personatges controlats per ordinador en jocs d'estratègia, com ara agents d'IA en escacs o jocs de rol.

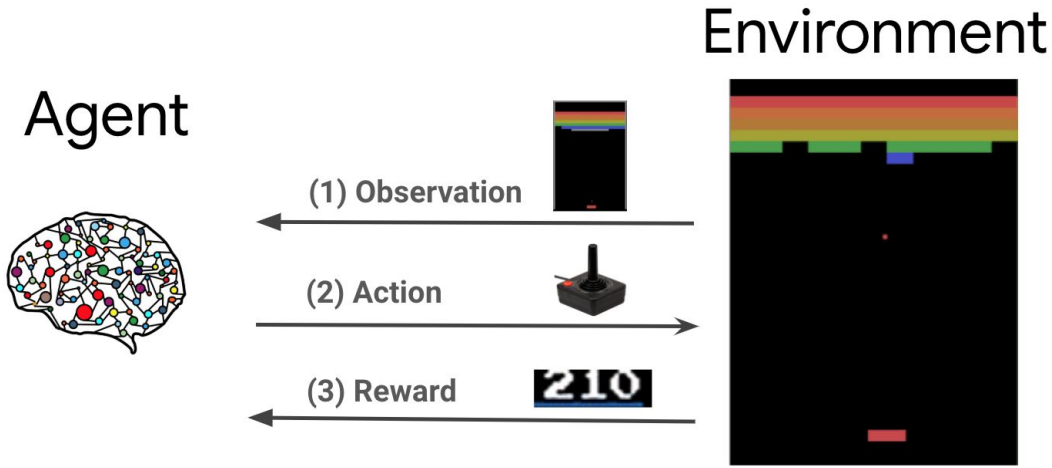
Agents de xarxes de transport intel·ligent:

- Vehicles autònoms que prenen decisions de conducció en temps real basades en les condicions de trànsit, normatives de seguretat i preferències de l'usuari.

Agents d'Optimització de Recursos:

- Sistemes que optimitzen l'ús d'energia, aigua o altres recursos en infraestructures intel·ligents, com ara edificis o xarxes elèctriques.

Exemples de casos d'ús



Exemples de casos d'ús

Exemples de casos d'ús

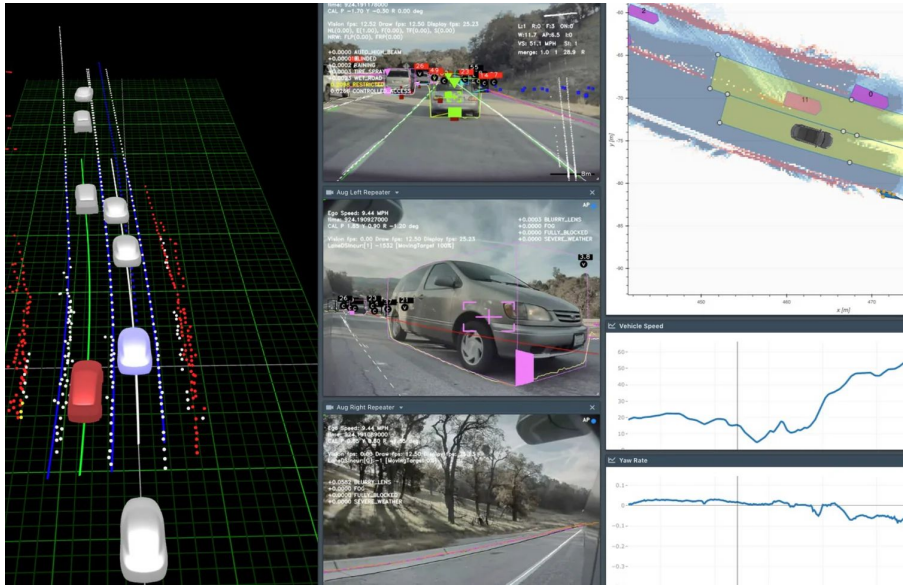
Entorn: definició

L'entorn en un model de RL és el context en què l'agent pren decisions i executa accions per aconseguir els seus objectius. L'entorn pot incloure tots els elements amb els quals l'agent interactua, com ara objectes físics, altres agents, condicions ambientals, dades d'entrada, i qualsevol altre factor rellevant per a l'aplicació en qüestió.

Entorn: característiques

- **Dinàmic**: L'entorn pot canviar de manera imprevisible o en resposta a les accions de l'agent.
- **Estocàstic**: Les reaccions de l'entorn poden tenir una certa dosi d'aleatorietat o incertesa.
- **Observabilitat**: L'agent pot tenir o no accés complet a la informació sobre l'estat de l'entorn.
- **Determinisme**: En alguns casos, l'entorn pot ser totalment previsible, mentre que en altres pot ser inherentment imprevisible.
- **Continuïtat**: L'entorn pot ser discret o contínu, depenent de la naturalesa de les variables que el defineixen.

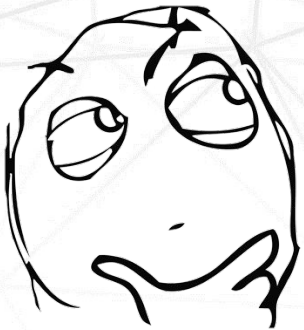
Entorn: exemples



vedrunapro
Vall Terrassa

Entorn: exemples

Com es defineix un entorn?



vedrunapro
Vall Terrassa

Definint un entorn

I. Identificar els Components Clau:
   A. Comença identificant els components essencials de l'entorn que seran rellevants per a l'aplicació específica. Això pot incloure objectes físics, variables d'estat, recompenses i altres factors rellevants.

II. Descriure l'Estadístic:
   A. Defineix l'estat de l'entorn. L'estat és una representació de tota la informació pertinent sobre l'entorn en un moment determinat. Això pot incloure la posició dels objectes, les condicions ambientals, l'estat del sistema, etc.

Definint un entorn 2

I. Identificar les Accions Possibles:
   A. Determina les accions que l'agent pot prendre en l'entorn. Aquestes accions poden ser moviments físics, decisions de control, eleccions d'estratègies, etc.

II. Definir les Regles de Transició:
   A. Estableix com l'entorn canvia d'un estat a un altre en resposta a les accions de l'agent. Aquestes regles descriuen com l'entorn reacciona a les accions de l'agent i com evoluciona amb el temps.

Definint un entorn 3

I. Especificar el Model de Recompenses:
   A. Defineix el sistema de recompenses que s'utilitzarà per guiar l'aprenentatge de l'agent. Això implica determinar quan i com s'assignen recompenses positives o negatives en funció de les accions de l'agent i l'estat de l'entorn.

II. 6. Considerar la Observabilitat:
   A. Decideix si l'agent té accés complet o parcial a la informació sobre l'estat de l'entorn. Això pot afectar les estratègies d'aprenentatge i la presa de decisions de l'agent.

III. 7. Validar i Ajustar:
   A. Prova el teu entorn per assegurar-te que les regles i les dinàmiques funcionen com es desitja. Pots ajustar i iterar en la definició de l'entorn segons les necessitats i els resultats de les proves.

Estats

- Un estat és una representació de les condicions o situació actual de l'entorn en un moment específic. És la informació que l'agent utilitza per prendre decisions i seleccionar accions.
- Els estats poden ser representats de manera contínua o discreta. En estats continus, els valors poden ser qualsevol nombre real en un interval determinat, mentre que en estats discrets, els valors són específics i finits.

Exemples d’estats

- Tres en ratlla: En aquest joc, els estats poden ser les diferents configuracions del taulell, incloent-hi les posicions de les fitxes (X o O).
- Comerç Financer: En un entorn de comerç, els estats podrien ser els diferents valors dels indicadors financers, com ara preus d'accions, volum de transaccions, etc.

Recompensa

- La recompensa és una senyal o valor que l'agent rep de l'entorn en resposta a les accions que pren. La recompensa reflecteix l'èxit o el fracàs de l'agent en aconseguir els seus objectius i pot ser positiva o negativa.
- La recompensa defineix l'objectiu de l'agent en l'entorn. L'agent cerca maximitzar la recompensa acumulada al llarg del temps.
- Les recompenses poden ser immediates, quan es reben immediatament després de l'acció de l'agent, o retardades, quan es reben en un moment posterior.

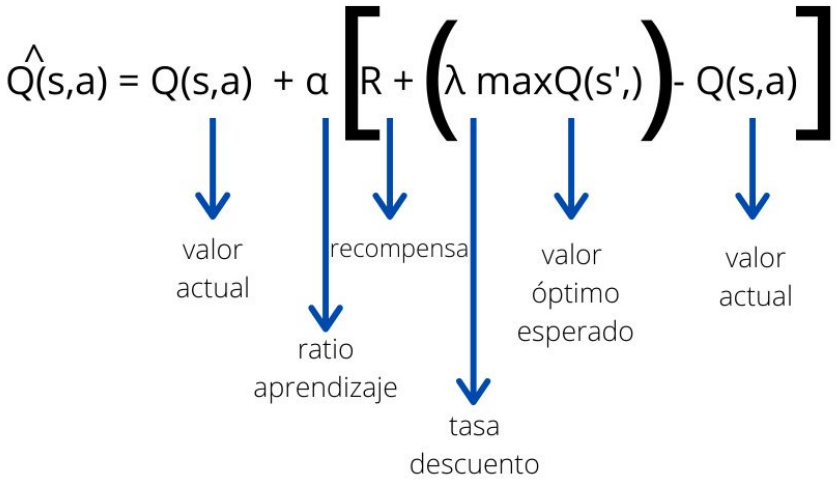
Recompensa

- La recompensa és el principal mecanisme de realimentació per a l'agent en RL. Guia l'aprenentatge i ajuda a l'agent a ajustar les seves estratègies per maximitzar la recompensa acumulada al llarg del temps.
- Una correcta definició de la funció de recompensa és crítica per al bon funcionament de l'algorisme d'aprenentatge per reforç i pot influir significativament en el rendiment de l'agent en l'entorn.

Q-learning

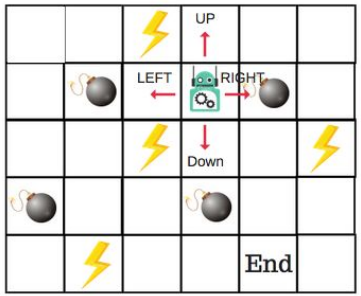
Q-learning és un algorisme d'aprenentatge per reforç (RL) que s'utilitza per aprendre una política òptima d'actuació en un entorn desconegut. És un mètode de RL basat en valor que apren directament la funció Q, que és una funció que assigna valors a parelles estat-acció i indica la recompensa esperada d'executar una acció en un determinat estat i seguir una determinada política d'actuació.

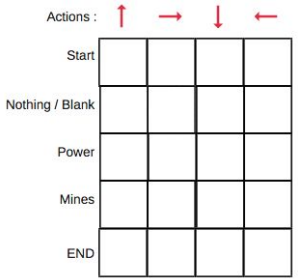
La funció Q


$\hat{Q}(s,a) = Q(s,a) + \alpha \left[R + \left(\lambda \max Q(s',) \right) - Q(s,a) \right]$
valor actual
ratio aprendizaje
recompensa
tasa descuento
valor óptimo esperado
valor actual

Q-table

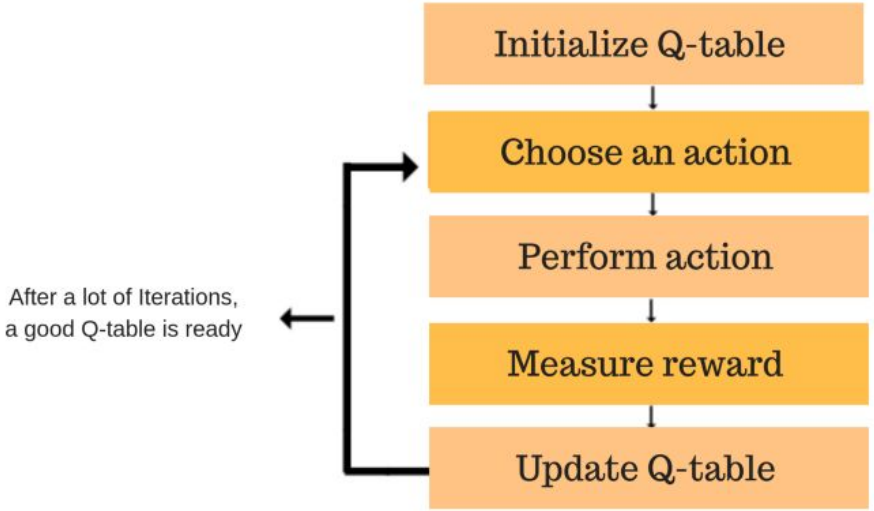
Q-Table és només un nom elegant per a una simple taula de cerca on calculem les màximes recompenses futures esperades per acció en cada estat. Bàsicament, aquesta taula ens guiarà a la millor acció en cada estat.





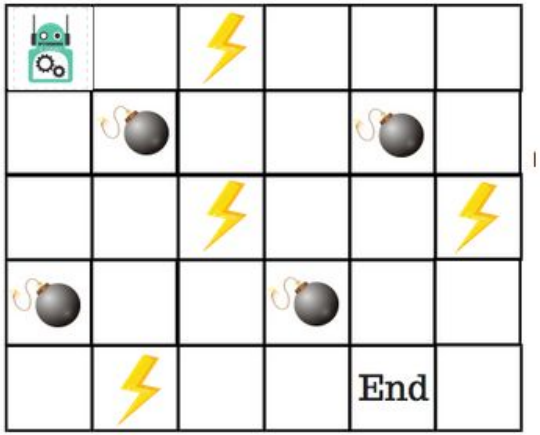
Cada puntuació de la Q-Table serà la màxima recompensa futura esperada que el robot rebrà si pren aquesta acció en aquest estat. Es tracta d'un procés iteratiu, ja que necessitem millorar la Q-Table en cada iteració.

Com entrenar?



Com entrenar?

Actions :	↑	→	↓	←
Start	0	0	0	0
Nothing / Blank	0	0	0	0
Power	0	0	0	0
Mines	0	0	0	0
END	0	0	0	0



Com entrenar?

Pas 2 i 3. Escollir i realitzar una acció i calcular la seva recompensa.

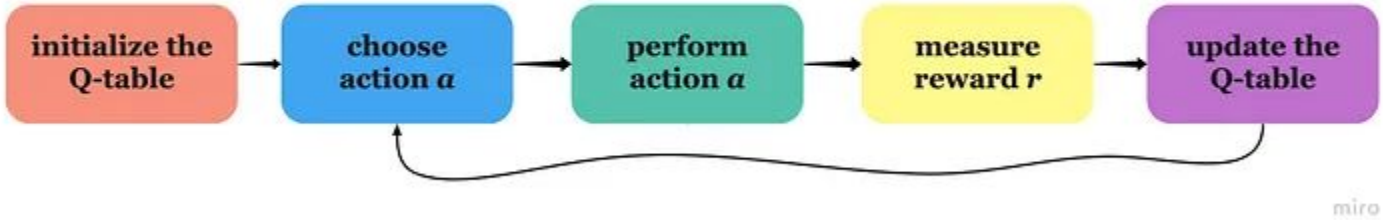
- Triarem una acció (a) en l'estat (s) basat en la Q-Table. Però, comhem mencionat anteriorment, quan l'episodi comença inicialment, cada valor de Q és 0. Així que ara entra en joc el concepte de compensació d'exploració i explotació.

Passos 4 i 5: avaluar

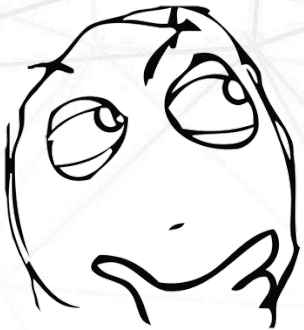
- Ara hem pres una acció i observat un resultat i recompensa.Necessitem actualitzar la funció Q(s,a).

Com entrenar?

Repetir el procés fins que finalitzin els episodis o tinguem algun sistema de terminació.



Preguntes?



vedrunapro
Vall Terrassa