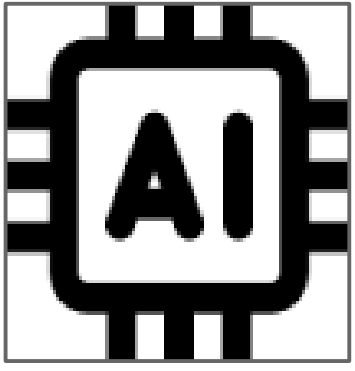
Introducción a la Inteligencia Artificial



Introducción al Aprendizaje Reforzado

Contacto: ia@frh.utn.edu.ar

En esta Presentación

1. Introducción al Aprendizaje Reforzado (RL)

- Una tercera categoría de ML
- El aprendizaje reforzado

2. Componentes de un sistema de RL

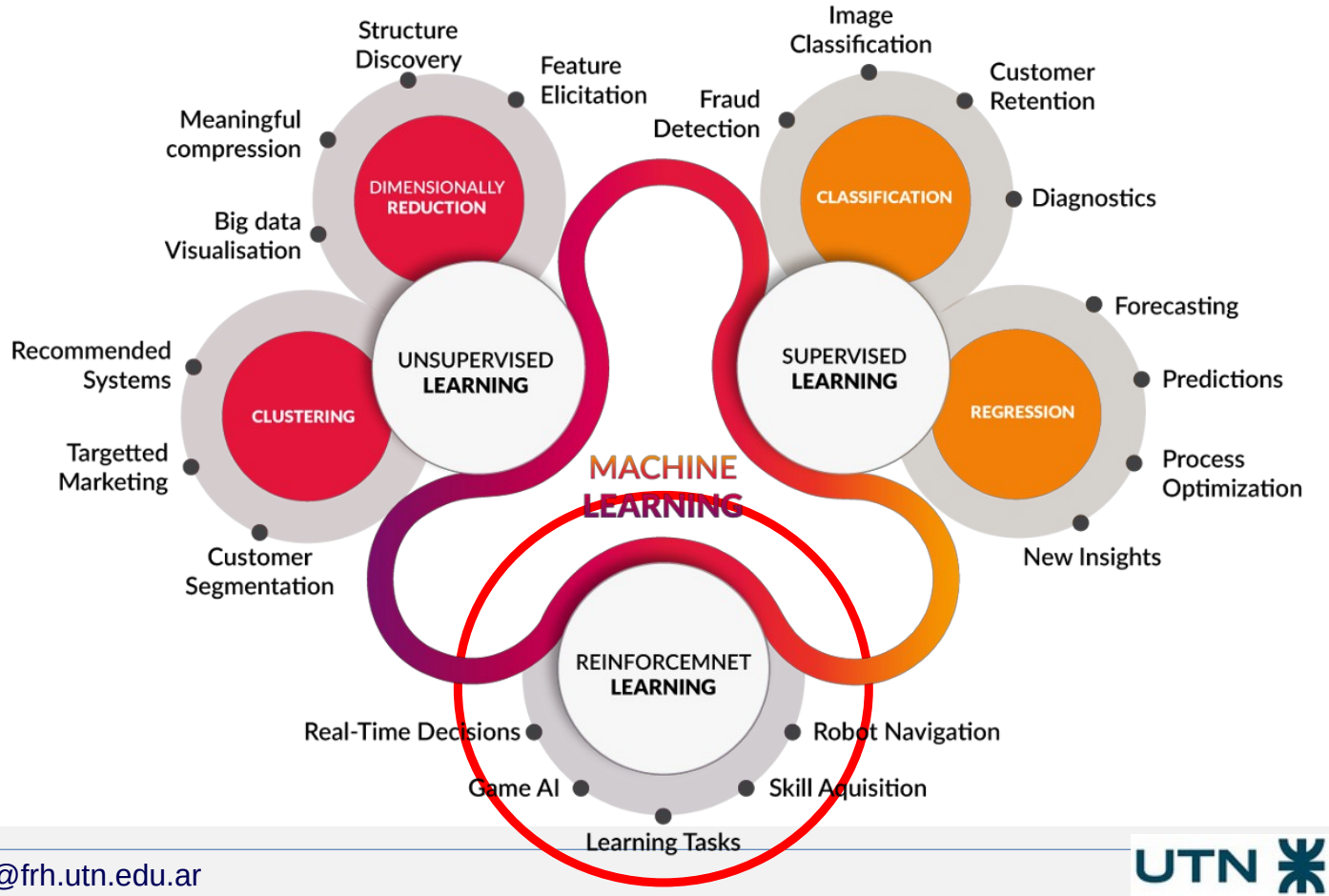
- Descripción de componentes

3. El ciclo del RL

4. Algunos Usos y aplicaciones de Aprendizaje Reforzado

5. Conclusiones

Introducción al Aprendizaje reforzado



Una tercera categoría de ML

Se diferencia del aprendizaje supervisado

El aprendizaje supervisado esta basado en un set de ejemplos que han sido etiquetados previamente.

Esto no es beneficioso en problemas interactivos. Esta técnica termina siendo inefectiva en territorio no explorado, donde el agente debe aprender de su propia experiencia [1].

Una tercera categoría de ML

Se diferencia del aprendizaje no-supervisado

El aprendizaje no supervisado se centra en encontrar estructuras dentro de una colección de datos, cuando el aprendizaje por refuerzo se enfoca en aumentar la señal de recompensa [1].

Una tercera categoría de ML

El aprendizaje reforzado se considera un tercer paradigma de aprendizaje automático.

El aprendizaje reforzado (RL)

Es una rama del aprendizaje automático en la cual *un agente aprende a tomar decisiones mediante la interacción con un entorno*.

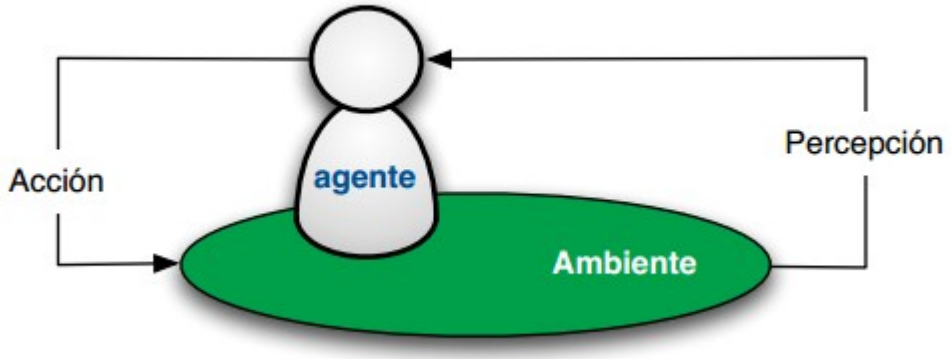


Imagen obtenida de [2]

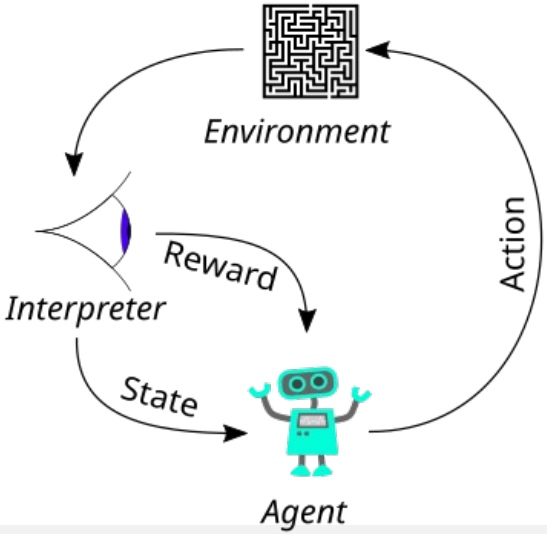
El aprendizaje reforzado (RL)

A través de un sistema de recompensas y castigos, el agente aprende qué acciones lo acercan a su objetivo y qué acciones no.

Este tipo de aprendizaje es útil cuando no se tiene acceso directo a las respuestas correctas, sino que el agente debe aprender por experiencia.

El aprendizaje reforzado (RL)

Un agente toma acciones en un entorno, que se interpreta en una recompensa y una representación del estado, que se retroalimentan al agente [3].



Componentes de un sistema RL

Agente: Entidad que toma decisiones y aprende.

Entorno: Espacio donde y con el cual el agente interactúa.

Acción: Las decisiones que el agente puede tomar.

Recompensa: Un valor de retroalimentación que indica qué tan buena fue la acción tomada por el agente.

Estado: Representación del entorno en un momento dado.

Componentes de un sistema RL

Política: La estrategia que sigue el agente para seleccionar acciones.

Función de valor: Representa la expectativa de recompensa futura desde un estado particular.

El agente y su entorno

Es una entidad que percibe y actúa sobre un entorno [4].

Basándose en esta definición, se pueden caracterizar distintos agentes de acuerdo con los atributos que posean y métodos que definen su comportamiento para resolver un determinado problema [2].

El agente y su entorno

Los agentes están compuestos de sensores y actuadores.

Un agente es cualquier cosa capaz de percibir su medio ambiente y esto lo logra con la ayuda de sensores, y es capaz de actuar en ese medio utilizando actuadores [2].

El agente y su entorno

En términos matemáticos se puede decir que el comportamiento del agente viene dado por la función del agente que proyecta una percepción dada en una acción [4].

Ejemplo: Una aspiratora automatica

La aspiradora puede percibir en que cuadrante se encuentra y si hay suciedad o no en él. Puede elegir si se mueve de cuadricula, aspirar la suciedad o no hacer nada.

Una función muy simple para el agente vendría dada por: si la cuadricula en la que se encuentra está sucia, entonces aspirar, de otra forma cambiar de cuadricula. La función del agente siempre se representa en una tabla [3].

Objetivos y recompensas

El propósito principal del agente es optimizar la cantidad total de recompensa que obtiene a lo largo de su interacción con el entorno.

Esto implica no solo maximizar la gratificación inmediata, sino también maximizar la recompensa acumulada en el plazo [5].

Objetivos y recompensas

En el contexto del aprendizaje por refuerzo, una característica fundamental es la utilización de una *señal de recompensa* para formalizar el concepto de objetivo [5].

Objetivos y recompensas

Ejemplo: el proceso de enseñar a un robot a caminar.

En este escenario, los investigadores han diseñado sistemas en los que el robot recibe una recompensa en cada paso de tiempo, la cual es proporcional al avance que logra en su trayectoria hacia adelante.

Este enfoque motiva al robot a perfeccionar su habilidad para caminar de manera efectiva. [5].

Objetivos y recompensas

Ejemplo: Aprender a jugar a las damas.

Para que un agente aprenda a jugar a las damas, las recompensas naturales son +1 al ganar, -1 al perder y 0 al empatar y para todas las posiciones no terminales [5].

Objetivos y recompensas

Precaución: La cuidadosa definición de recompensas es esencial para alinear los intereses de los agentes con nuestros objetivos, garantizando que trabajen en la dirección deseada sin que sea necesario especificar cada paso del proceso.

Conjunto de estados

Representa todas las posibles situaciones que puede experimentar un sistema.

Estos estados pueden ser estados físicos, emocionales o cualquier otro tipo de categorización que sea relevante para el problema en cuestión [4].

Política

Una política define el modo en que el agente se comporta en un momento definido, una política vendría a mapear que acciones se deberían llevar a cabo dado los estados que se han percibido del medio ambiente [1].

Política

Puede representar una función simple o una tabla de búsqueda.

En la mayoría de los casos puede involucrar un calculo extensivo.

La política representa el núcleo del agente y es suficiente para determinar el comportamiento de éste [1].

Función de valor

La mayoría de los algoritmos de aprendizaje por refuerzo se fundamentan en la estimación de funciones de valor.

Estas funciones se aplican a estados individuales o pares estado-acción, y su propósito es evaluar cuán beneficioso es para el agente encontrarse en un estado dado o que tan bueno es realizar una acción específica en un estado determinado [5].

Función de valor Optima

Resolver una tarea de aprendizaje por refuerzo significa, en términos generales, encontrar una política que logre muchas recompensas a largo plazo.

Para procesos de decisión finitos, se puede definir con precisión una política óptima [5].

Función de valor Optima

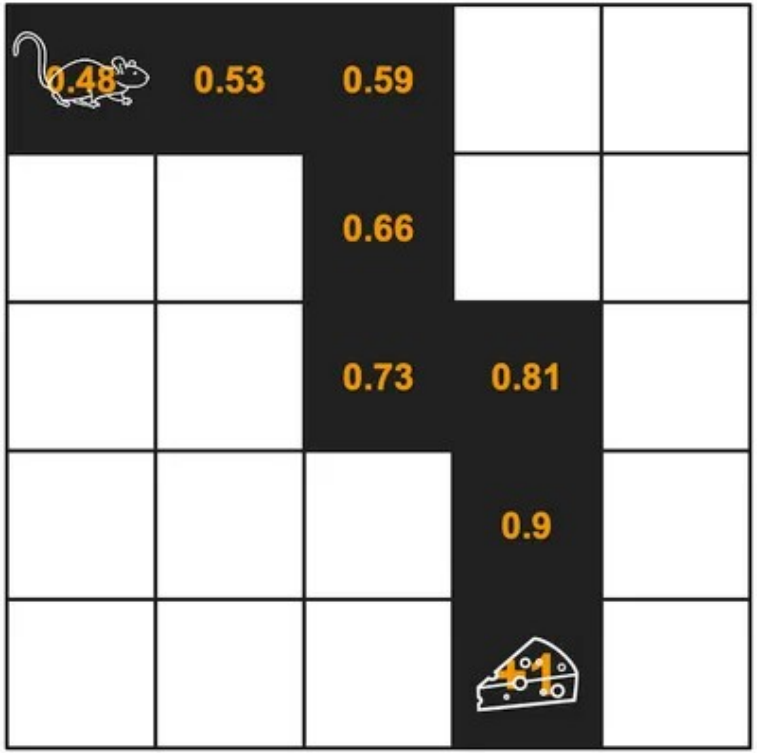


Imagen obtenida de [5].

Ciclo de Aprendizaje Reforzado

1)El agente observa el estado actual del entorno.

2)Toma una acción basada en su política.

3)El entorno cambia de estado y le proporciona una recompensa.

4)El agente ajusta su política para maximizar la recompensa futura.

Algunos ejemplos de aplicaciones

1) Autos autónomos

Coches Autónomos: Empresas como Tesla y Waymo usan RL en la formación de sus vehículos autónomos para navegar por diferentes entornos, tomar decisiones sobre cuándo frenar, acelerar o girar, e interactuar con el tráfico.

Los agentes de RL son entrenados para maximizar la seguridad y la eficiencia mediante la simulación de escenarios del mundo real.

Algunos ejemplos de aplicaciones

2) NLP (Procesamiento del Lenguaje Natural)

Sistemas de diálogo: Los chatbots pueden mejorar sus respuestas mediante RL. Por ejemplo, en un sistema de servicio al cliente, el chatbot aprende a generar respuestas más útiles y satisfactorias mediante la retroalimentación de las interacciones pasadas con los usuarios (si el cliente quedó satisfecho o no).

Algunos ejemplos de aplicaciones

3) Finanzas y Trading

Trading Algorítmico: En los mercados financieros, RL se usa para desarrollar agentes de trading que aprenden estrategias para comprar y vender activos financieros.

Gestión de carteras: También se usa RL para gestionar portafolios de inversión de forma dinámica. El agente aprende a redistribuir activos en una cartera en función de las fluctuaciones del mercado, maximizando el rendimiento a largo plazo.

Algunos ejemplos de aplicaciones

4) Salud y Medicina

Optimización de Tratamientos:

En medicina personalizada, el RL se ha utilizado para diseñar tratamientos personalizados, optimizando la dosis de medicamentos para cada paciente en función de su respuesta individual.

Un ejemplo es el ajuste de dosis de medicamentos en el tratamiento de enfermedades crónicas.

Conclusiones

Los agentes aprenden a tomar decisiones a través de la interacción con un entorno dinámico, recibiendo *recompensas o penalizaciones*.

Conclusiones

Los agentes aprenden a tomar decisiones a través de la interacción con un entorno dinámico, recibiendo *recompensas o penalizaciones*.

A diferencia del aprendizaje supervisado, donde los ejemplos etiquetados guían al modelo, el *aprendizaje reforzado permite que un agente aprenda mediante la prueba y el error*, lo que lo hace efectivo en problemas interactivos y en situaciones donde no se dispone de datos de antemano.

Conclusiones

El objetivo clave en *RL es la maximización de una señal de recompensa acumulada a largo plazo*. Busca guiar el comportamiento del agente hacia decisiones óptimas basadas en la retroalimentación recibida del entorno.

Se presentaron algunas *aplicaciones* de RL.

Referencias

[1] Silva, Miguel (2019). “Aprendizaje por Refuerzo: Introducción al mundo del RL”, Medium. En línea: https://medium.com/aprendizaje-por-refuerzo-introducci%C3%B3n-al-mundo-del/aprendizaje-por-refuerzo-introducci%C3%B3n-al-mundo-del-rl-1fcfbaa1c87

[2] Ceballos, karla (2014).”Inteligencia Artificial”. Portafolio Digital Espam-Mfl. En línea: https://inteligenciaartificialkarlacevallos.wordpress.com/2014/11/04/2-1-agentes-y-su-entorno/

[3] Wikipedia (2024). “Aprendizaje por refuerzo”. En línea: https://es.wikipedia.org/wiki/Aprendizaje_por_refuerzo

[4] Russell, S., Norvig, P. 2008. Inteligencia Artificial Un Enfoque Moderno. Segunda Edición. Pearson Education. España.

[5] Cerretani, Joan (2023). “Aprendizaje por refuerzo (RL)”. Medium. En línea: https://medium.com/@joancerretanids/aprendizaje-por-refuerzo-rl-cap%C3%ADtulo-2-introducci%C3%B3n-parte-2-recompensas-retornos-y-markov-36ee763ea9bf