



TIC



Actividad 5

Aprendizaje por Refuerzo (RL)

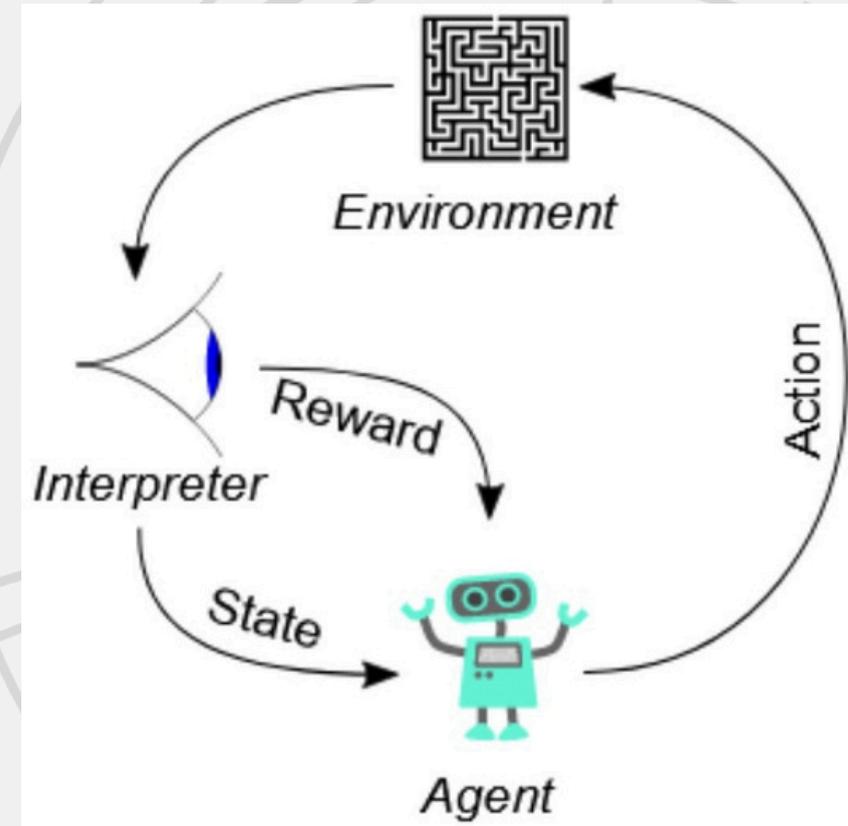
Aprendizaje por Refuerzo (RL)



TIC



El Aprendizaje por Refuerzo emerge como un paradigma distintivo en el campo del aprendizaje automático, presentando una perspectiva única en comparación con los enfoques supervisado y no supervisado. En lugar de depender de conjuntos de datos etiquetados o no etiquetados, el RL se centra en aprender a través de la interacción activa con un entorno dinámico. Esta interacción implica la toma de decisiones secuenciales, donde un agente busca maximizar una señal de recompensa a lo largo del tiempo.





TIC



Aprendizaje por Refuerzo (RL)

Características Clave del Aprendizaje por Refuerzo:



Interacción con un Entorno



Secuencialidad en la Toma de Decisiones



Recompensas y Aprendizaje a Largo Plazo

El Aprendizaje por Refuerzo destaca la importancia de la interacción secuencial en la toma de decisiones, proporcionando un marco valioso para abordar problemas complejos en situaciones dinámicas. Su aplicabilidad en una variedad de campos demuestra su relevancia en la resolución de problemas del mundo real que involucran decisiones secuenciales y adaptativas.

Relevancia del Aprendizaje por Refuerzo en Toma de Decisiones Secuenciales:



Aplicaciones en Inteligencia Artificial



Exploración y Explotación



Toma de Decisiones Contextualizada



Aprendizaje a Partir de la Experiencia



Exploración y Explotación:

El aprendizaje por refuerzo aborda el equilibrio entre la exploración de nuevas acciones y la explotación de acciones conocidas, lo que es esencial en entornos dinámicos y cambiantes.



Exploración y Explotación:

El aprendizaje por refuerzo aborda el equilibrio entre la exploración de nuevas acciones y la explotación de acciones conocidas, lo que es esencial en entornos dinámicos y cambiantes.



Aplicaciones en Inteligencia Artificial

El RL se destaca en situaciones donde la toma de decisiones es secuencial y adaptativa, como en juegos, robótica, conducción autónoma y asistentes virtuales.



Secuencialidad en la Toma de Decisiones

En lugar de tomar decisiones independientes, el RL se enfoca en la toma de decisiones secuenciales a lo largo del tiempo. Cada acción del agente afecta el estado futuro del entorno y, por lo tanto, las recompensas acumulativas a lo largo de una secuencia de acciones.



Secuencialidad en la Toma de Decisiones

En lugar de tomar decisiones independientes, el RL se enfoca en la toma de decisiones secuenciales a lo largo del tiempo. Cada acción del agente afecta el estado futuro del entorno y, por lo tanto, las recompensas acumulativas a lo largo de una secuencia de acciones.



Interacción con un Entorno

A diferencia del aprendizaje supervisado, donde se proporcionan ejemplos etiquetados, y del no supervisado, que explora estructuras latentes en datos no etiquetados, el RL implica que un agente interactúe con un entorno en constante cambio.



Aprendizaje a Partir de la Experiencia

El agente en RL aprende a través de la experiencia y la retroalimentación del entorno, lo que lo hace adecuado para problemas donde la información es parcial o incompleta.

Conceptos Básicos Aprendizaje por Refuerzo



TIC

El Aprendizaje por Refuerzo se fundamenta en varios conceptos clave que definen la dinámica entre un agente y su entorno. Estos conceptos proporcionan la base para entender cómo el agente toma decisiones secuenciales para maximizar las recompensas acumulativas a lo largo del tiempo.



Agente



Entorno



Acciones



Estados



Recompensas

Estos conceptos constituyen la base del aprendizaje por refuerzo, permitiendo a los agentes aprender estrategias efectivas para interactuar con su entorno y lograr objetivos específicos.



Estados

Los estados representan la información relevante sobre el entorno en un momento dado. La toma de decisiones del agente se basa en la percepción y comprensión de su estado actual. Los estados pueden ser variables como la posición de un robot, la duración del mercado financiero o la disposición de las piezas en un juego de mesa.



Acciones

Las acciones son las decisiones que toma el agente en respuesta a su entorno. Estas pueden ser movimientos físicos, elecciones estratégicas o cualquier otra interacción que afecte al entorno. Ejemplos de acciones incluyen movimientos de un robot, decisiones de inversión financiera o selección de jugadas en un juego.



Entorno

El entorno representa el contexto en el que opera el agente. Es el sistema que responde a las acciones del agente y proporciona retroalimentación en forma de recompensas. Los entornos pueden variar desde juegos virtuales hasta entornos físicos como fábricas o vehículos autónomos.



Agente

El agente en el aprendizaje por refuerzo es la entidad que toma decisiones y realiza acciones dentro de un entorno específico. Su objetivo es aprender una estrategia que maximice las recompensas a lo largo del tiempo. Ejemplos de agentes incluyen robots autónomos, asistentes virtuales y sistemas de control industrial.



Recompensas

Las recompensas son señales de retroalimentación que el entorno proporciona al agente en respuesta a sus acciones. El objetivo del agente es maximizar las recompensas acumulativas a lo largo del tiempo. Las recompensas pueden ser positivas, negativas o neutrales, indicando la calidad de las acciones tomadas.

Algoritmos de Aprendizaje por Refuerzo



TIC

Existen varios algoritmos diseñados para que un agente aprenda a tomar decisiones secuenciales para maximizar las recompensas en un entorno. Aquí se presentan algunos de los algoritmos fundamentales:

Métodos de Valoración:



Iteración de Valor



Evaluación de Políticas

Métodos de Optimización de Políticas:



Monte Carlo



Métodos de Gradiente

Aprendizaje Profundo y DQN:



Aprendizaje Profundo (Deep RL)



Deep Q-Network (DQN)

Estos algoritmos son fundamentales para abordar problemas de RL en diversas aplicaciones, desde juegos hasta robótica y gestión de recursos. La elección del algoritmo depende del problema específico y de las características del entorno en el que opera el agente.



Métodos de Gradiente

Estos algoritmos buscan directamente la política que maximiza la recompensa esperada. Utilizan el gradiente de la política para realizar ajustes.



Deep Q-Network (DQN)

Un algoritmo específico de Deep RL que utiliza una red neuronal para aproximar la función Q. DQN ha demostrado éxito en juegos y entornos complejos.



Iteración de Valor

Este enfoque implica iterativamente evaluar y mejorar la función de valor de cada estado. Se actualizan los valores de los estados basándose en las recompensas y las estimaciones anteriores.



Evaluación de Políticas

Se evalúan y mejoran directamente las políticas del agente. La idea es ajustar la probabilidad de tomar ciertas acciones en determinados estados para mejorar el rendimiento global.



Monte Carlo

Este método se basa en la idea de estimar la recompensa total esperada para cada estado mediante la simulación de episodios completos. Luego, se ajustan las políticas para maximizar las recompensas esperadas.



Aprendizaje Profundo (Deep RL):

Aplicar redes neuronales profundas al Aprendizaje por Refuerzo. Esto permite manejar estados y acciones de alta dimensionalidad.



TIC



▶ TALENTO
TECH

