TIC

Módulo 1

Lección 4

Conceptos del Aprendizaje por Refuerzo

Universidad Tecnológica de Pereira

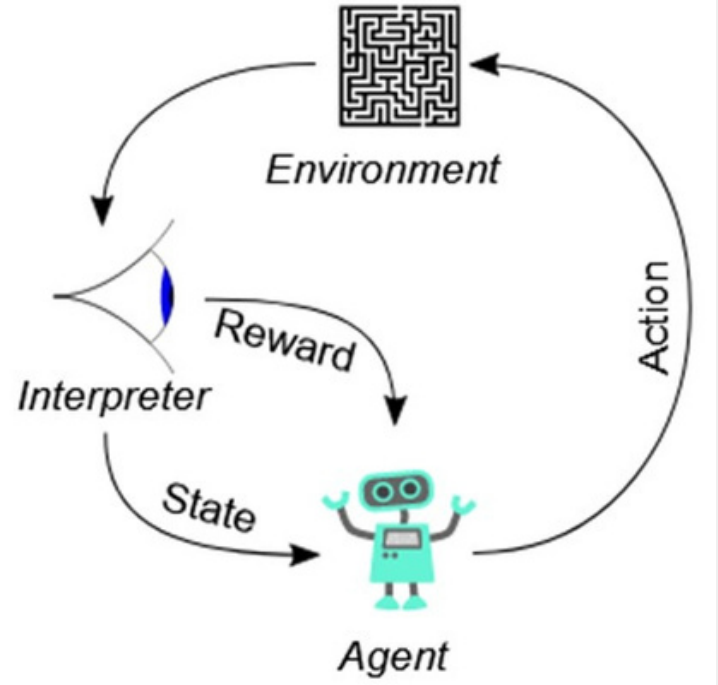
faceIT

Contenido

Haz clic sobre los títulos para navegar en cada tema.

Tema 1. Introducción

El Aprendizaje por Refuerzo (RL), 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.



Volver a Contenido

Características Clave del Aprendizaje por Refuerzo

Interacción con un entorno	Secuencialidad en la toma de decisiones	Recompensas y aprendizaje a largo plazo
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 interactúa con un entorno en constante cambio.	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.	La guía principal para el agente en RL es la señal de recompensa, que indica la calidad de sus acciones. El objetivo es aprender políticas y estrategias que maximicen las recompensas acumulativas a lo largo del tiempo, fomentando un aprendizaje a largo plazo.

Relevancia del aprendizaje por refuerzo en la toma de decisiones secuenciales

1. Aplicaciones en Inteligencia Artificial: se destaca en situaciones donde la toma de decisiones es secuencial y adaptativa como en juegos, robótica, conducción autónoma y asistentes virtuales.

2. Exploración y Explotación: aborda el equilibrio entre la exploración de nuevas acciones y la explotación de acciones conocidas, lo que es esencial en entornos dinámicos y cambiantes.

3. Toma de Decisiones Contextualizada: permite que un agente tome decisiones basadas en el contexto actual del entorno, ajustándose a condiciones cambiantes y mejorando la adaptabilidad.

4. Aprendizaje a Partir de la Experiencia: el agente 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.

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.

Tema 2. Conceptos básicos aprendizaje por refuerzo

El Aprendizaje por Refuerzo (RL), 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.

Volver a Contenido

Dinámina Agente–Entorno

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.

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.

Conceptos importantes

Acciones	Estados	Recompensas
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.	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 configuración del mercado financiero o la disposición de las piezas en un juego de mesa.	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.

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.

Tema 3. Proceso de Decisión de Markov (MDP)

El Proceso de Decisión de Markov (MDP), es un modelo matemático fundamental en el Aprendizaje por Refuerzo (RL), que describe la interacción entre un agente y su entorno a lo largo del tiempo. Su principal característica es la propiedad de Markov, que establece que el futuro es independiente del pasado dado el estado presente.

Volver a Contenido

Explicación y elementos del MDP

Estados (S): el conjunto de todos los posibles estados en los que puede encontrarse el entorno. En el contexto del MDP, la propiedad de Markov implica que la información contenida en el estado actual es suficiente para predecir el siguiente estado.

Acciones (A): el conjunto de todas las posibles acciones que puede tomar el agente en un estado dado, las acciones afectan la transición entre estados y, por lo tanto, el resultado final.

Probabilidades de Transición (P): define la probabilidad de pasar de un estado a otro, dado que una acción específica representa la dinámica del entorno y cómo las acciones del agente afectan las transiciones entre estados.

Explicación y elementos del MDP

Recompensas (R): la recompensa es una función que asigna un valor numérico a cada par estado-acción. Indica la utilidad o deseabilidad de tomar una acción específica en un estado dado.

Política (
π): la política es una estrategia que el agente sigue para tomar decisiones.

Puede ser determinista o estocástica y define la probabilidad de tomar cada acción en cada estado.

Representación Gráfica

Un MDP, se puede representar gráficamente mediante un grafo dirigido donde los nodos representan estados y las aristas representan transiciones posibles entre estados bajo ciertas acciones. Se pueden agregar etiquetas a las aristas para indicar las probabilidades de transición y las recompensas asociadas.

Representación Gráfica

La comprensión del MDP es esencial para diseñar políticas efectivas y algoritmos de aprendizaje por refuerzo. Este modelo proporciona una base sólida para abordar problemas de toma de decisiones secuenciales en una variedad de aplicaciones, desde robótica hasta gestión de recursos y juegos.

Tema 4. Algoritmos de Aprendizaje por Refuerzo

En el campo del Aprendizaje por Refuerzo (RL), existen varios algoritmos diseñados para que un agente aprenda a tomar decisiones secuenciales para maximizar las recompensas en un entorno.

En este tema te presentaremos algunos de los algoritmos más importantes.

Volver a Contenido

Métodos de valoración

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.

Métodos de optimización de políticas

IteraMonte 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.

Métodos de Gradiente: estos algoritmos buscan directamente la política que maximiza la recompensa esperada; utilizan el gradiente de la política para realizar ajustes.

Aprendizaje profundo y DQN

Aprendizaje Profundo (Deep RL): aplicar redes neuronales profundas al Aprendizaje por Refuerzo, esto permite manejar estados y acciones de alta dimensionalidad.

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.

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.