

Qué es un agente en el aprendizaje por reforzamiento

José Carlos Machicao

Hoy les voy a contar sobre el agente computacional que se construye para implementar un modelo de Aprendizaje por Reforzamiento (RL). Imaginemos que este agente es una entidad autónoma, como un pequeño explorador inteligente. Su misión: aprender a tomar las mejores decisiones en un entorno del que, al principio, no sabe absolutamente nada.

Lo primero que el agente aprende es que hay dos formas de actuar:

- Puede explorar, es decir, probar acciones nuevas para descubrir cómo funciona el entorno.
- puede explotar lo que ya sabe para obtener los mejores resultados posibles.

Es como cuando ustedes juegan un videojuego nuevo: al principio prueban botones para ver qué pasa (explorar), pero después de varias partidas ya saben qué estrategia les da puntos (explotar).

Nuestro agente tiene algo muy importante: memoria. Puede recordar sus experiencias, aunque a veces también es bueno que olvide para no aferrarse a decisiones equivocadas. Sin embargo, al principio no entiende qué es lo que debe recordar. Solo sabe que hay algo llamado recompensa: un valor que recibe cada vez que interactúa con el entorno. Puede ser alto o bajo, pero al inicio no tiene ni idea de qué significa ni cómo conseguirlo.

Entonces empieza su fase de exploración. Hace muchas pruebas, algunos experimentos son exitosos y otros no. Poco a poco, el agente empieza a notar patrones: 'Si hago esta acción en este estado, la recompensa mejora'. Es aquí donde surge una especie de vocabulario interno:

- Define estados, que son descripciones de su relación con el entorno (por ejemplo, a qué distancia está de un objetivo).
- También tiene un vocabulario de acciones, que son las decisiones que puede tomar.

A medida que sigue explorando, el agente empieza a conectar estos estados y acciones con las recompensas que obtiene. Esa conexión cada vez se vuelve más clara y estructurada. Al final, decide organizar toda esa información en una tabla de correspondencia entre estados, acciones y recompensas: su famosa Q-table.

La Q-table se convierte en su fuente principal de conocimiento. Ahora, en lugar de pensar demasiado en sus acciones y estados, consulta directamente su tabla: '¿Qué decisión me da la mejor recompensa aquí?' Esto le permite actuar de manera mucho más eficiente.

Sin embargo, hay algo más interesante: el agente nunca deja de aprender. Si nota que su recompensa se ha estabilizado o que podría estar mejor, entra en una fase de reflexión:

- '¿Será que me estoy perdiendo algo?'
- '¿Debería volver a explorar?'

Y entonces decide salir de su zona de confort para buscar nuevas formas de mejorar. Porque un agente inteligente no solo aprende rápido, sino que también sabe cuándo cuestionar lo que ya aprendió.

En resumen, este agente empieza sin saber nada, pero a través de exploración, memoria y aprendizaje constante, construye un modelo cada vez más estructurado y eficiente para tomar decisiones. Es la esencia del Aprendizaje por Reforzamiento.