

Aprendizaje por refuerzo

Resumen: métodos tabulares

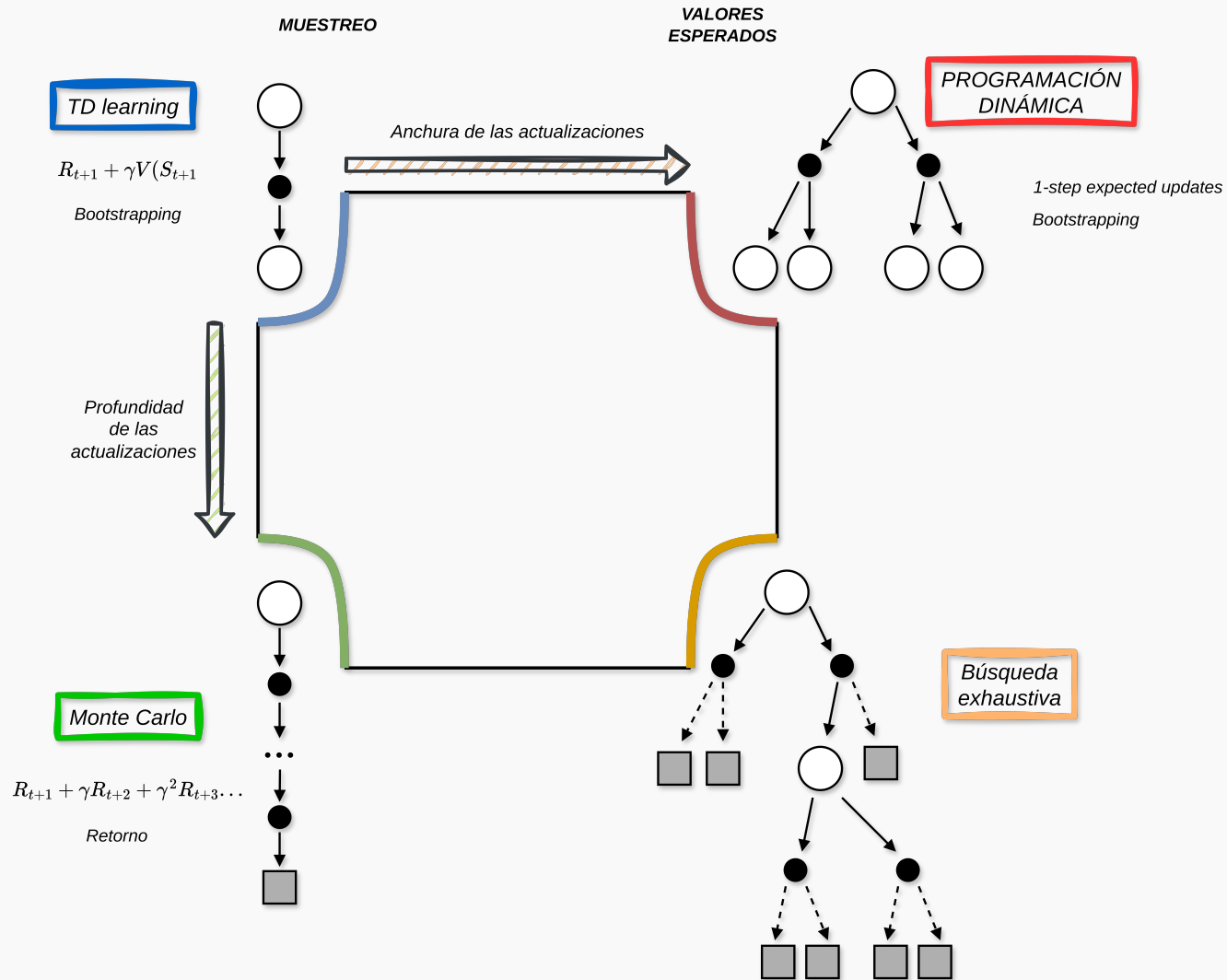
Antonio Manjavacas Lucas

manjavacas@ugr.es

1. Visión general
2. Conceptos clave
3. Limitaciones

Visión general

Resumen



Conceptos clave

¿Diferencia entre problemas **episódicos** y **continuados**?

¿Diferencia entre problemas **episódicos** y **continuados**?

- **Episódicos**: problemas divisibles en episodios con una duración determinada, desde un estado inicial hasta un estado terminal.
- **Continuados**: no existen estados terminales.

¿Diferencia entre **estado** y **observación**?

¿Diferencia entre **estado** y **observación**?

- **Estado**: contiene información **completa** sobre el estado actual del entorno.
- **Observación**: contiene la información **parcial** percibida por el agente. Es un **subconjunto** de la información contenida en el estado.

Diferencia entre MDP y POMDP.

¿Diferencia entre **valor** y **recompensa**?

¿Diferencia entre **valor** y **recompensa**?

- **Valor**: retorno esperado a partir de un estado o par acción-estado.
- **Recompensa**: valor inmediato percibido al alcanzar un estado o realizar una determinada acción desde un estado.

¿Qué es el **retorno**?

¿Qué es el **retorno**?

Retorno: recompensa acumulada al final de un episodio o secuencia de *time steps*.

¿Diferencia entre **exploración** y **explotación**?

¿Diferencia entre **exploración** y **explotación**?

- **Exploración**: elección de acciones subóptimas que pueden conducir a nuevas experiencias/transiciones/recompensas no percibidas.
- **Explotación**: aplicación de la política (óptima) actual para maximizar una función de recompensa.

¿Diferencia entre **predicción** y **control**?

¿Diferencia entre **predicción** y **control**?

- **Predicción**: estimación de las funciones de valor para una política dada. Consiste en **evaluar** una política.
- **Control**: encontrar la política óptima que maximice la recompensa acumulada. Implica **evaluación** y **mejora** de la política actual.

¿Diferencia entre métodos *on-policy* y *off-policy*?

¿Diferencia entre métodos *on-policy* y *off-policy*?

- ***On-policy***: la política empleada para generar experiencia y aprender es **la misma** que se emplea para actuar.
- ***Off-policy***: una política genera experiencia (**comportamiento**) mientras que otra se emplea para actuar (**objetivo**).

¿Diferencia entre *model-based* y *model-free*?

¿Diferencia entre *model-based* y *model-free*?

- **Model-based**: métodos de RL basados en el aprendizaje y/o aprovechamiento de modelos para generar experiencia y **planificar** un comportamiento óptimo.
- **Model-free**: métodos de RL que no requieren un modelo del entorno y **aprenden directamente** a partir de interacción real.

Conceptos clave

Programación dinámica

- Evaluación de la política.
- Mejora de la política.
- Iteración de la política.
- DP síncrona vs. asíncrona.
- Iteración de la política generalizada (GPI).

Bandits

- Exploración vs. explotación.
- ϵ -greedy.
- *Upper Confidence Bound*
- *Thompson Sampling*
- Actualizaciones incrementales

Planificación

- Dyna-Q
- Dyna-Q+
- *MC tree search*

Métodos basados en muestreo

- Monte Carlo.
- Inicios de exploración.
- *Importance sampling.*
- *On-policy vs. off-policy.*
- *TD learning.*
- SARSA.
- *Q-learning.*
- *Expected SARSA.*
- Métodos *n-step.*

Limitaciones

Métodos tabulares

Los métodos vistos hasta el momento se denominan **tabulares** debido a la forma en que almacenan y gestionan la información.

Por ejemplo, en el caso de las **políticas**, estas pueden representarse como una **tabla** que relaciona estados y acciones, o estados, acciones y valores.

- Ej. Q-table.

*Asumimos que el número de estados y acciones es **discreto y limitado**, por lo que su gestión es computacionalmente viable.*

Limitaciones

? ¿Pero qué ocurre si el **espacio de estados** / **acciones** es infinito?

? ¿Pero qué ocurre si el **espacio de estados** / **acciones** es infinito?

... o lo suficientemente grande como para ser **inabarcable** por los métodos vistos 🤔

? ¿Pero qué ocurre si el **espacio de estados** / **acciones** es infinito?

... o lo suficientemente grande como para ser **inabarcable** por los métodos vistos 🤔

¡Lo veremos en la siguiente parte! 🧠 😊

Aprendizaje por refuerzo

Resumen: métodos tabulares

Antonio Manjavacas Lucas

manjavacas@ugr.es