

# Introducción a aprendizaje por refuerzo

## Sesión 1

Gabriel Muñoz  
Máster en Inteligencia Artificial, 2019-2020

## Sobre mí

---

Gabriel Muñoz

*Lead Data Scientist en Intelygenz (<https://intelygenz.com/>)*

Co-fundador de *MAD\_RL*, comunidad de aprendizaje por refuerzo con base en Madrid

*[gabrielenrique.munoz@campusviu.es](mailto:gabrielenrique.munoz@campusviu.es)  
<https://es.linkedin.com/in/gabrielmunozrios>*



## Sobre la asignatura

---

El contenido de la asignatura se organizará en cuatro bloques básicos:

- En el primer bloque empezaremos con un vistazo general al estado del arte y a los retos que nos encontramos con estas técnicas de inteligencia artificial.
- Seguidamente nos meteremos más en detalle con conceptos y terminología necesaria para poder entender las posibilidades de esta rama.
- Pasaremos luego a dos de los algoritmos que más han impactado y que son la base de métodos más avanzados: *Deep Q-Networks* y *Policy Gradients*.
- Terminaremos con una sesión práctica que servirá como piedra de arranque para la entrega.



## Sobre la asignatura

---

Respecto a los requisitos de la asignatura:

- Los conceptos teóricos que veremos se basarán en conceptos matemáticos y estadísticos. Hablaremos de distribuciones de probabilidad, de búsqueda óptima, de cadenas de *Markov*, etc. El conocer estos conceptos no es una restricción para el curso, pero si que es recomendable.
- En la parte práctica trabajaremos con Python y algunas librerías típicas de entornos para trabajar con datos, como por ejemplo *Keras* y *Tensorflow*. En la sesión práctica listaremos las librerías y diferentes opciones de instalación.
- En ambas partes será imprescindible conocer y tener cierta experiencia con modelos *Deep Learning* y, en concreto, con redes convolucionales.



## Sobre la asignatura

---

Para la evaluación se tendrá en cuenta una prueba final en la que se reflejará el conocimiento adquirido durante las sesiones.

Esta evaluación tendrá dos partes:

- Una parte teórica donde se propondrán preguntas breves para repasar los conceptos vistos.
- Una parte práctica sobre un reto propuesto, donde desarrollaremos parte de los algoritmos que veremos durante las sesiones.



# Índice

---

¿Qué es aprendizaje por refuerzo?

Estado del arte

Distintos enfoques de aprendizaje por refuerzo

Retos actuales y de futuro

Vista general

Conclusiones

Bibliografía recomendada



## ¿Qué es aprendizaje por refuerzo?

Estado del arte

Distintos enfoques de aprendizaje por refuerzo

Retos actuales y de futuro

Vista general

Conclusiones

Bibliografía recomendada



## ¿Qué es aprendizaje por refuerzo?

---

*“Aprender a partir de la interacción es una idea fundamental y común a todas las teorías del aprendizaje y de la inteligencia”*

*“Aprendizaje por refuerzo [...] es a la vez un problema, un conjunto de soluciones que funcionan bien sobre un conjunto de problemas y el campo que estudia esos problemas y los métodos que pueden solucionarlo.”*

Reinforcement learning: An Introduction, R. Sutton & A. Barto

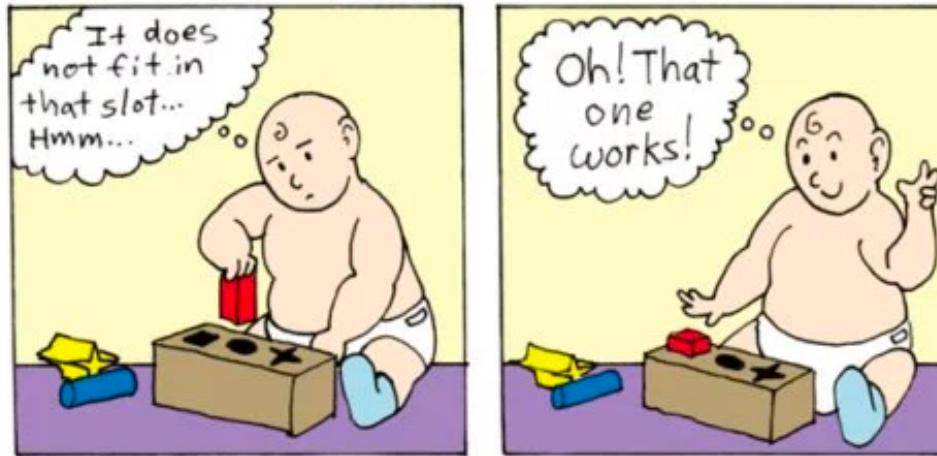
*“El objetivo del aprendizaje por refuerzo es entrenar un agente inteligente que es capaz de interactuar con un entorno de manera inteligente.”*

Deep Q Network vs Policy gradients, Felix Yu



## ¿Qué es aprendizaje por refuerzo?

Al final podemos ver el aprendizaje por refuerzo como el estudio y diseño de **agentes** que aprenden por medio de **prueba-y-error**.



## ¿Qué es aprendizaje por refuerzo?

---

Si nos paramos y analizamos otras ramas dentro del aprendizaje basado en datos encontramos dos grandes conjuntos: métodos supervisados y métodos no supervisados. Podríamos añadir aprendizaje por refuerzo como un tercer conjunto dentro de esta clasificación.

Desde un punto de vista de datos, podemos hacer una relación de cada conjunto con el conocimiento que se puede obtener a partir de los datos:

**Análisis descriptivo → Métodos no supervisados**

**Análisis predictivo → Métodos supervisados**

**Análisis prescriptivo → Métodos aprendizaje por refuerzo**



## ¿Qué es aprendizaje por refuerzo?

---

Si nos paramos y analizamos otras ramas dentro del aprendizaje basado en datos encontramos dos grandes conjuntos: métodos supervisados y métodos no supervisados. Podríamos añadir aprendizaje por refuerzo como un tercer conjunto dentro de esta clasificación.

Desde un punto de vista de datos, podemos hacer una relación de cada conjunto con el conocimiento que se puede obtener a partir de los datos:

**Análisis descriptivo → Métodos no supervisados**

Análisis predictivo → Métodos supervisados

Análisis prescriptivo → Métodos aprendizaje por refuerzo



## ¿Qué es aprendizaje por refuerzo?

---

Si nos paramos y analizamos otras ramas dentro del aprendizaje basado en datos encontramos dos grandes conjuntos: métodos supervisados y métodos no supervisados. Podríamos añadir aprendizaje por refuerzo como un tercer conjunto dentro de esta clasificación.

Desde un punto de vista de datos, podemos hacer una relación de cada conjunto con el conocimiento que se puede obtener a partir de los datos:

Análisis descriptivo → Métodos no supervisados

**Análisis predictivo → Métodos supervisados**

Análisis prescriptivo → Métodos aprendizaje por refuerzo



## ¿Qué es aprendizaje por refuerzo?

---

Si nos paramos y analizamos otras ramas dentro del aprendizaje basado en datos encontramos dos grandes conjuntos: métodos supervisados y métodos no supervisados. Podríamos añadir aprendizaje por refuerzo como un tercer conjunto dentro de esta clasificación.

Desde un punto de vista de datos, podemos hacer una relación de cada conjunto con el conocimiento que se puede obtener a partir de los datos:

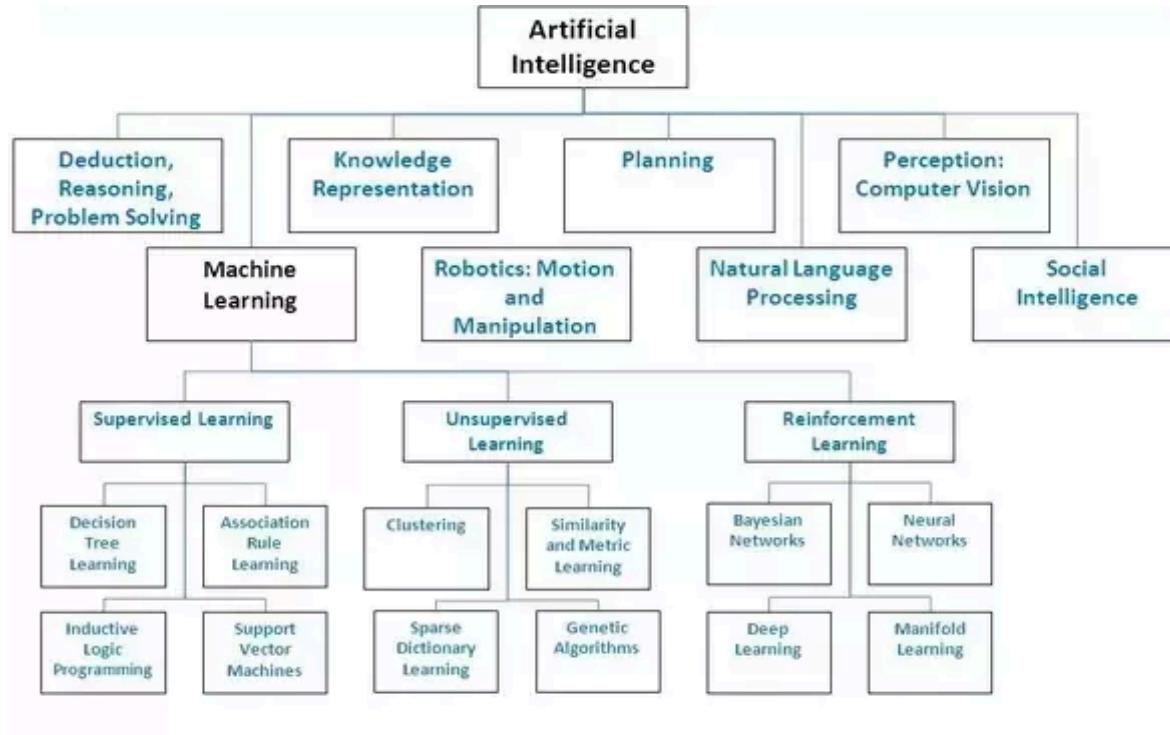
Análisis descriptivo → Métodos no supervisados

Análisis predictivo → Métodos supervisados

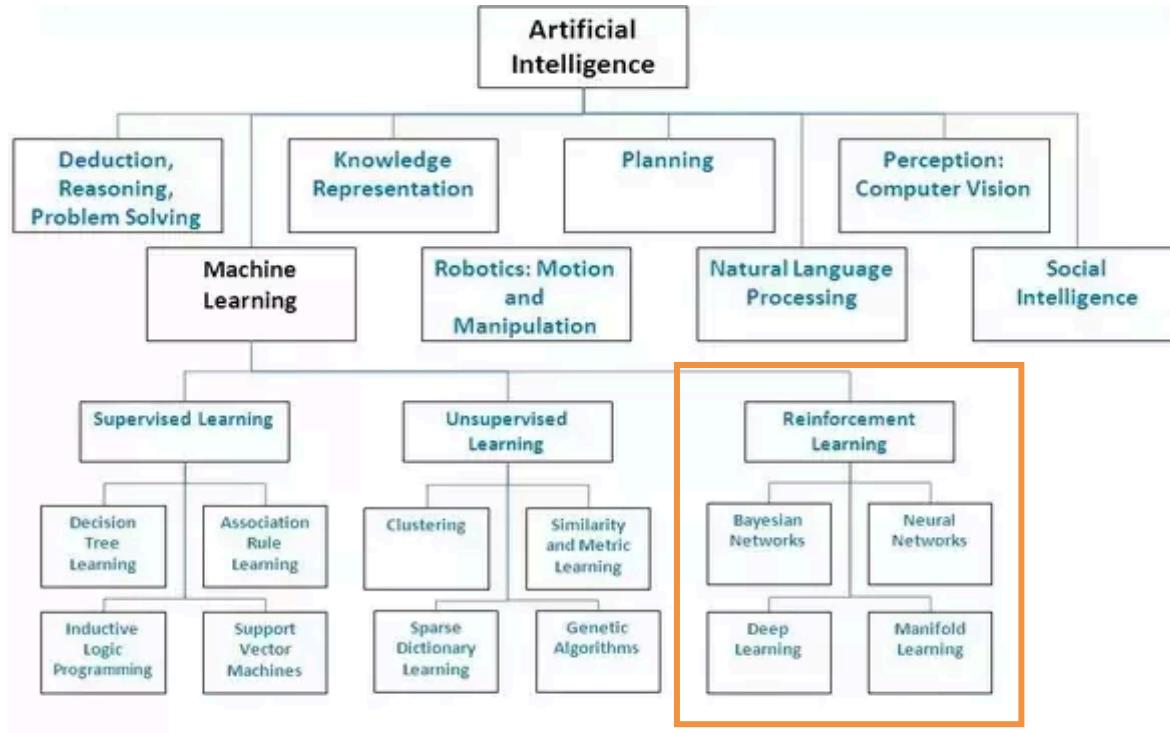
**Análisis prescriptivo → Métodos aprendizaje por refuerzo**



# ¿Qué es aprendizaje por refuerzo?



# ¿Qué es aprendizaje por refuerzo?



## ¿Qué es aprendizaje por refuerzo?

---

En mi opinión, el aprendizaje por refuerzo desarrolla de una manera natural e intuitiva lo que conocemos (o interpretamos) como aprendizaje.

Al producirse el aprendizaje de una manera parecida a la de los humanos, se podría empezar a hablar y discutir en términos de adaptación y/o transferencia de conocimiento entre distintos problemas. Precisamente la transferencia de conocimiento entre distintos problemas (*transfer learning*) es una rama de investigación con mucha fuerza hoy en día dentro de la inteligencia artificial y, en concreto, del *Deep Learning*.

Otro punto muy interesante es el concepto de bias o sesgo, que tanto impacto tiene en los datos con los que trabajamos. Dentro de una solución guiada por aprendizaje por refuerzo podríamos mitigar mucho este problema, ya que el aprendizaje se produce *from scratch*.



# Índice

---

¿Qué es aprendizaje por refuerzo?

**Estado del arte**

Distintos enfoques de aprendizaje por refuerzo

Retos actuales y de futuro

Vista general

Conclusiones

Bibliografía recomendada



## Estado del arte

---

Como en otras ramas de la inteligencia artificial, los primeros algoritmos y soluciones basadas en aprendizaje por refuerzo datan de hace algunas décadas. En nuestro caso, fue en los años 80/ principios de los 90 cuando hubo una tendencia positiva en el estudio y uso de algoritmos basados en aprendizaje por refuerzo.

Aunque los resultados fueron prometedores, debido a las capacidades computacionales y al desarrollo de otros algoritmos con enfoques similares en aquel momento (algoritmos evolutivos), las soluciones basadas en aprendizaje por refuerzo disminuyeron hasta hace unos años.

Como apunte, la principal diferencia entre algoritmos de aprendizaje por refuerzo y algoritmos evolutivos es que los agentes de algoritmos evolutivos no aprenden de la interacción con su entorno. Las reglas de su interacción son inamovibles, por lo que no pueden aprovechar todo la información de la que disponen para aprender y adaptarse al problema que intentan resolver.



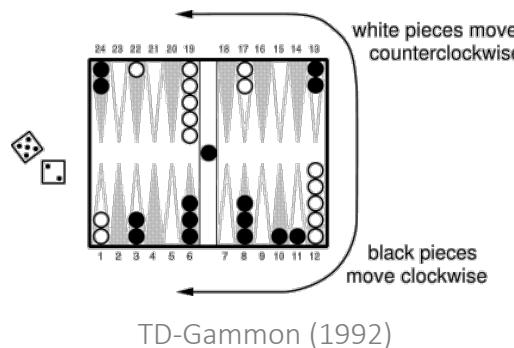
## Estado del arte

Como decíamos, hasta hace algunos años no hubo una explosión en este tipo de algoritmos y del potencial que tienen. Una relación directa ha sido el magnífico desarrollo de las técnicas de *Deep Learning* así como la capacidad computacional que ahora tenemos a nuestra disposición.

Igualmente, si tuviéramos que elegir tres hitos que sentaron las bases para hablar de aprendizaje por refuerzo hoy serían:



Minsky PHD thesis (1954)



TD-Gammon (1992)



DeepMind (2015)

## Estado del arte

---

Vamos a centrarnos en Google Deepmind. El gran hito que consiguió Google Deepmind con su solución fue que combinó *Deep Learning* con Aprendizaje por refuerzo.

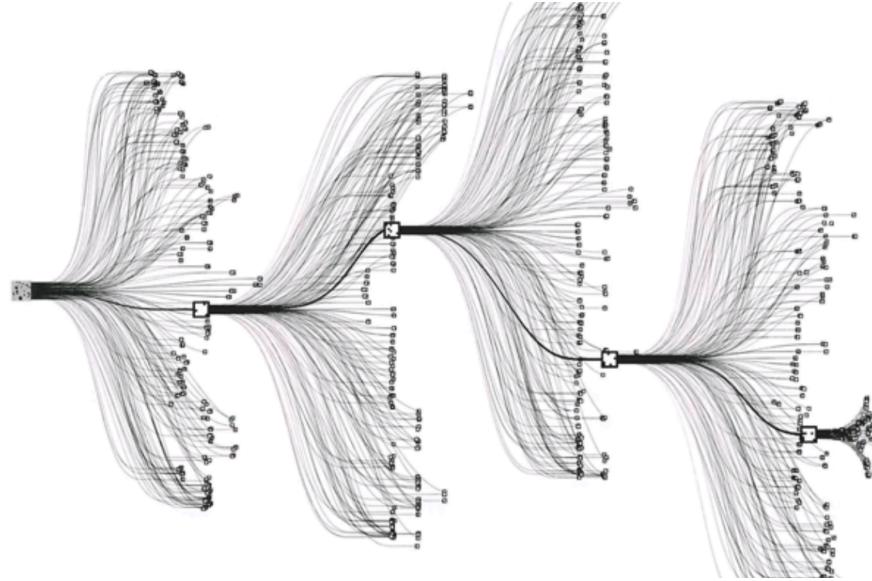
Hasta ese momento, las variables que se usaban en algoritmos por refuerzo eran variables *tabuladas* o definidas manualmente. Con la solución de DeepMind, se usó una red neuronal como función aproximadora para que el agente tomara decisiones, algo inaudito hasta el momento y con una base de inteligencia artificial formidable.

Este primer hito se realizó sobre un entorno de simulación de videojuegos de Atari.



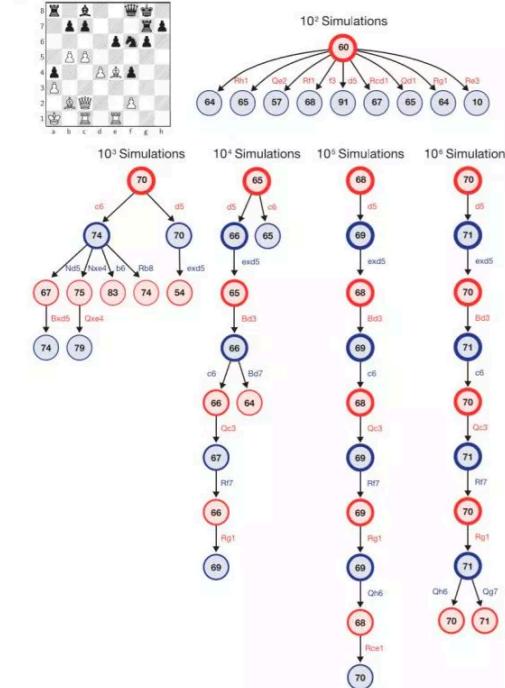
# Estado del arte

## Hitos de DeepMind: AlphaGo



# Estado del arte

## Hitos de DeepMind: AlphaZero



## Estado del arte

---

Otra compañía que ha sido un pulmón en el desarrollo del aprendizaje por refuerzo estos últimos años ha sido OpenAI.

La misión de OpenAI es la creación de una inteligencia artificial de carácter general con una fuerte base ética y de principios.

OpenAI han sido los responsables de algunos de los algoritmos que ahora mismo son estado del arte, así como de ser la primera compañía en ganar con inteligencia artificial en entornos de colaboración (y muy avanzados en cuanto a complejidad). Un ejemplo es el videojuego *DOTA2*.



## Estado del arte

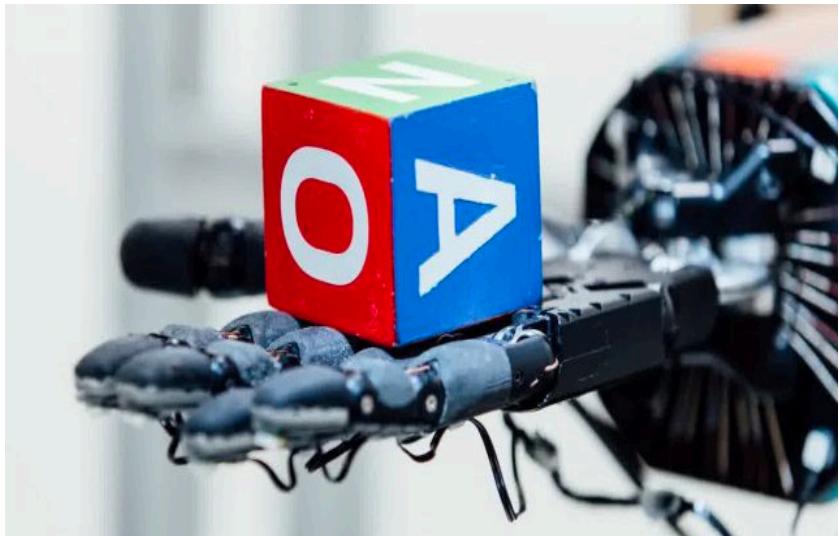
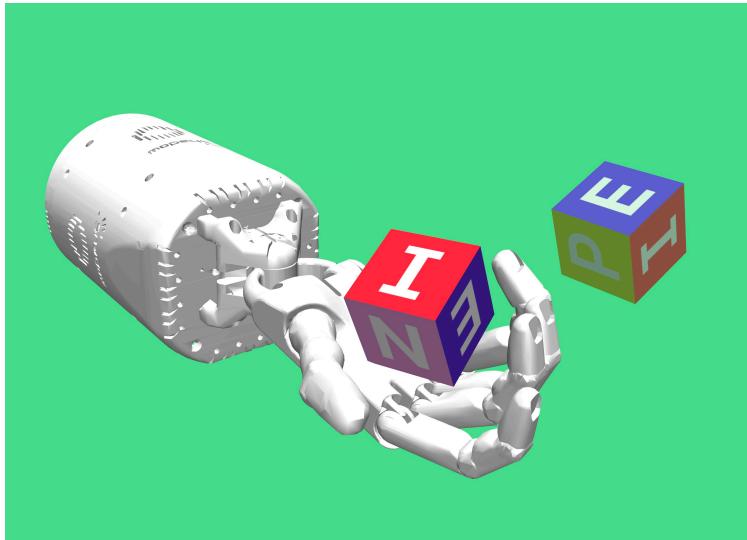
### Hitos de OpenAI: DOTA 2



## Estado del arte

---

### Hitos de OpenAI: Robótica



¿Qué es aprendizaje por refuerzo?

Estado del arte

## Distintos enfoques de aprendizaje por refuerzo

Retos actuales y de futuro

Vista general

Conclusiones

Bibliografía recomendada



## **Distintos enfoques del aprendizaje por refuerzo**

---

Una posible clasificación que podemos hacer cuando hablamos de aprendizaje por refuerzo está relacionado con el entorno donde se ejecutará la solución desarrollada.

En este sentido tenemos tres grupos principales:

**Entorno puramente digital - Ámbito de investigación**

**Entorno Digital-a-Real - Ámbito de investigación**

**Entorno puramente real - Ámbito de investigación y aplicación**

Vamos a entrar en detalle con cada enfoque.



## Distintos enfoques del aprendizaje por refuerzo



### Entorno puramente digital - Ámbito de investigación

Este es el entorno típico que nos encontraremos en la mayoría de ejemplos a día de hoy.

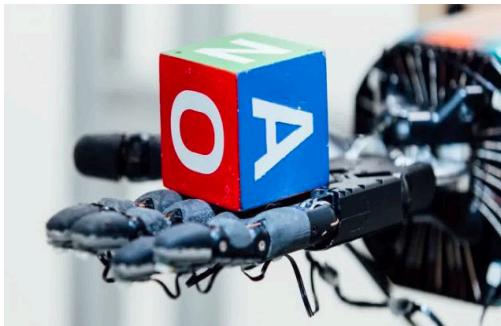
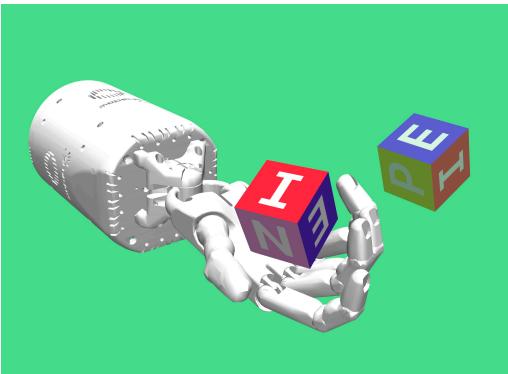


Trabajaremos en un entorno totalmente simulado y controlado para poder probar nuevos algoritmos, nuevos modelos, etc.

Este enfoque es fundamental para contrastar hipótesis teóricas y comprobar su posible aplicación en el mundo real.

## Distintos enfoques del aprendizaje por refuerzo

---



### Entorno Digital-a-Real - Ámbito de investigación

En este caso tenemos una situación parecida al enfoque anterior aunque ahora usaremos el conocimiento extraído en un entorno real.

Un caso común es la robótica, en el que se diseña toda la simulación en un entorno digital para usar la solución obtenida en brazos robóticos reales.

## Distintos enfoques del aprendizaje por refuerzo

---

### Entorno puramente real - Ámbito de investigación y aplicación

Este es el caso en el que nuestra solución se desplegará al servicio de los usuarios, comúnmente llamado *producción*.

También es la situación en la que el impacto que nuestra solución tiene es más tangible, ya que las soluciones alcanzadas son aplicables en el día a día.

Algunos ejemplos de soluciones son: Optimización de rutas, coches autónomos, modelos de optimización para control de procesos, etc.

Es común englobar en este grupo de soluciones de aprendizaje por refuerzo junto a algunos tipos de modelos generativos (GANs).



# Índice

---

¿Qué es aprendizaje por refuerzo?

Estado del arte

Distintos enfoques de aprendizaje por refuerzo

**Retos actuales y de futuro**

Vista general

Conclusiones

Bibliografía recomendada



## Retos actuales y de futuro

---

El principal reto en el que el aprendizaje por refuerzo se centra actualmente es encontrar casos de uso en los que se pueda aplicar de una manera beneficiosa.

Es verdad que en los últimos años se han ido desarrollando soluciones con aplicaciones reales en ámbitos como la robótica, control automático, etc. Pero todavía no tienen un impacto de negocio como para tomar a estas soluciones más en “serio”.

Para la mayoría de estos casos hay soluciones tradicionales que funcionan muy bien y no necesitan de la complejidad que necesita el aprendizaje por refuerzo.



## Retos actuales y de futuro

---

Una de las complejidades que presentan las soluciones basadas en aprendizaje por refuerzo, y que veremos en las siguientes sesiones, es la cantidad de hiperparámetros que se necesitan configurar para que una solución funcione.

Y no sólo eso, ¿qué significa que una *solución funcione*?

Veremos en siguientes sesiones conceptos para poder discutir sobre esta pregunta, pero lo que está claro es que el concepto de que algo sea útil o no sigue estando muy sesgado a los ojos de quien lo mire (como es normal por otra parte).



## Retos actuales y de futuro

---

Otro reto es que aún teniendo mucho dominio de la materia y conociendo muy bien todos los conceptos, el control sobre el proceso de aprendizaje es vago.

En otras disciplinas de la inteligencia artificial, como por ejemplo en la aplicación de técnicas de Deep Learning, tenemos una metodología que, aunque sea empírica, podemos controlar y medir cómo el sistema evoluciona y si lo está haciendo de forma favorable o no.

En el caso del aprendizaje por refuerzo no se dan las mismas circunstancias. Dependiendo del algoritmo, del reto y de la información disponible, la forma de medir la bondad de una solución difiere y por tanto le añade una complejidad extra a su diseño.



## Retos actuales y de futuro

---

Aún así, todos los grupos de investigación comparten la idea de que el potencial que tiene este enfoque de la inteligencia artificial es *infinito*.

La mayoría de problemas que ahora mismo podemos solucionar son problemas más cercanos a la automatización que al razonamiento y adaptación que una inteligencia real permitiría.

Es aquí donde también se presenta un gran reto: ser capaz de encontrar la madurez suficiente para afrontar verdaderos problemas de Inteligencia Artificial siguiendo esta metodología.



# Índice

---

¿Qué es aprendizaje por refuerzo?

Estado del arte

Distintos enfoques de aprendizaje por refuerzo

Retos actuales y de futuro

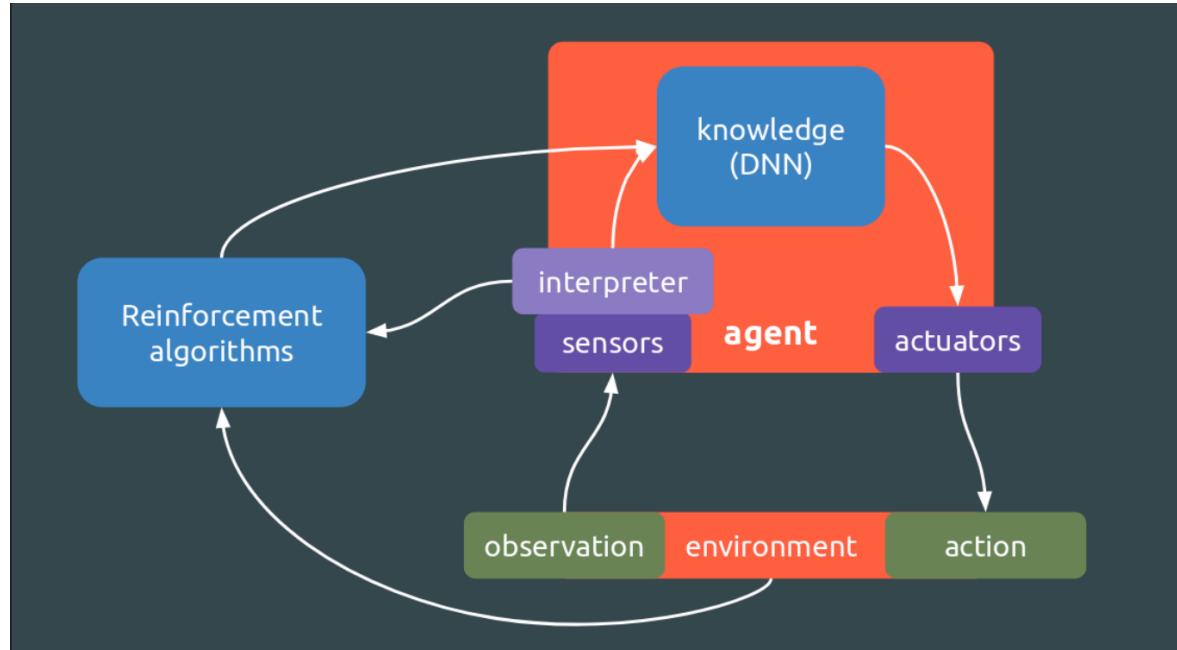
**Vista general**

Conclusiones

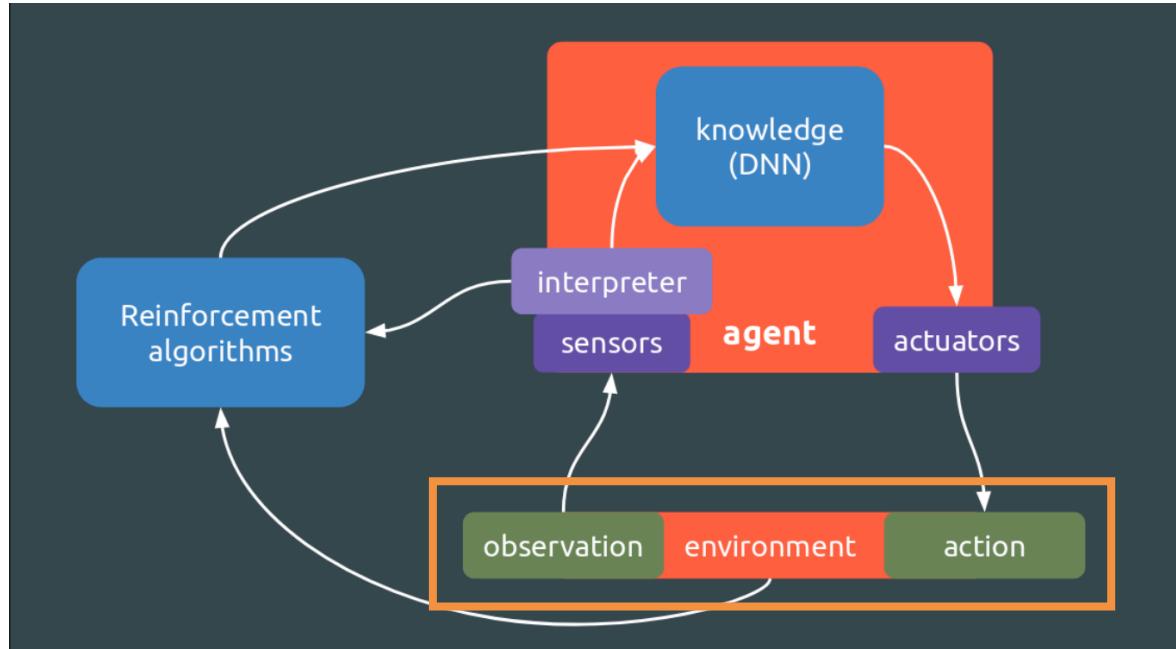
Bibliografía recomendada



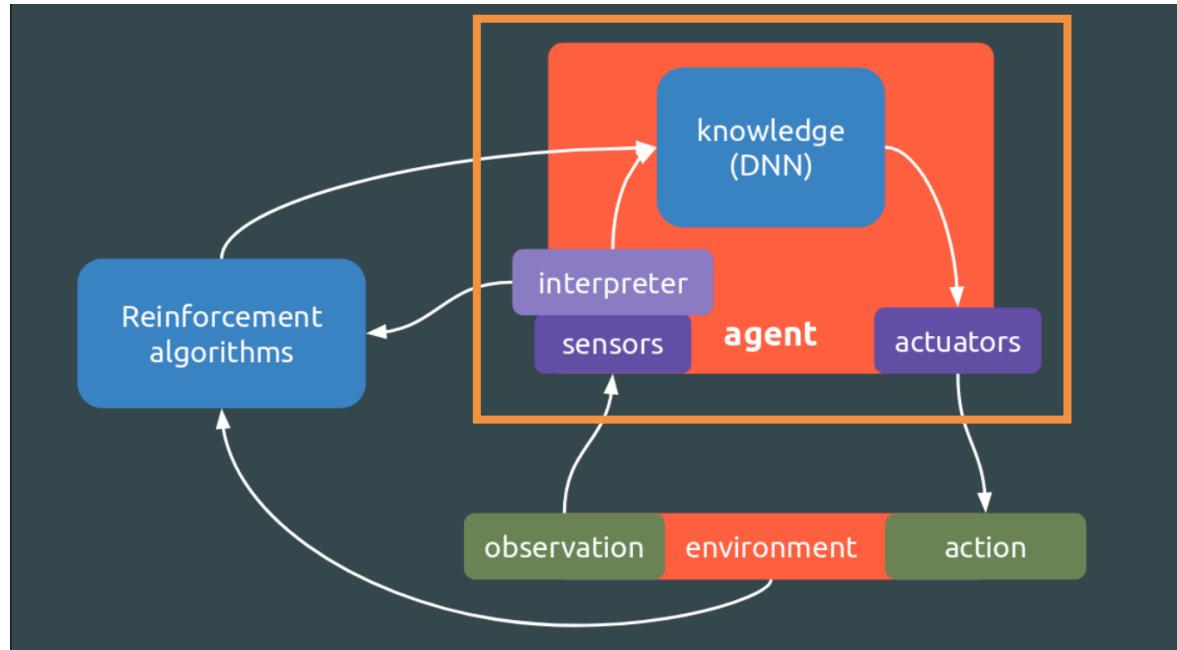
# Vista general



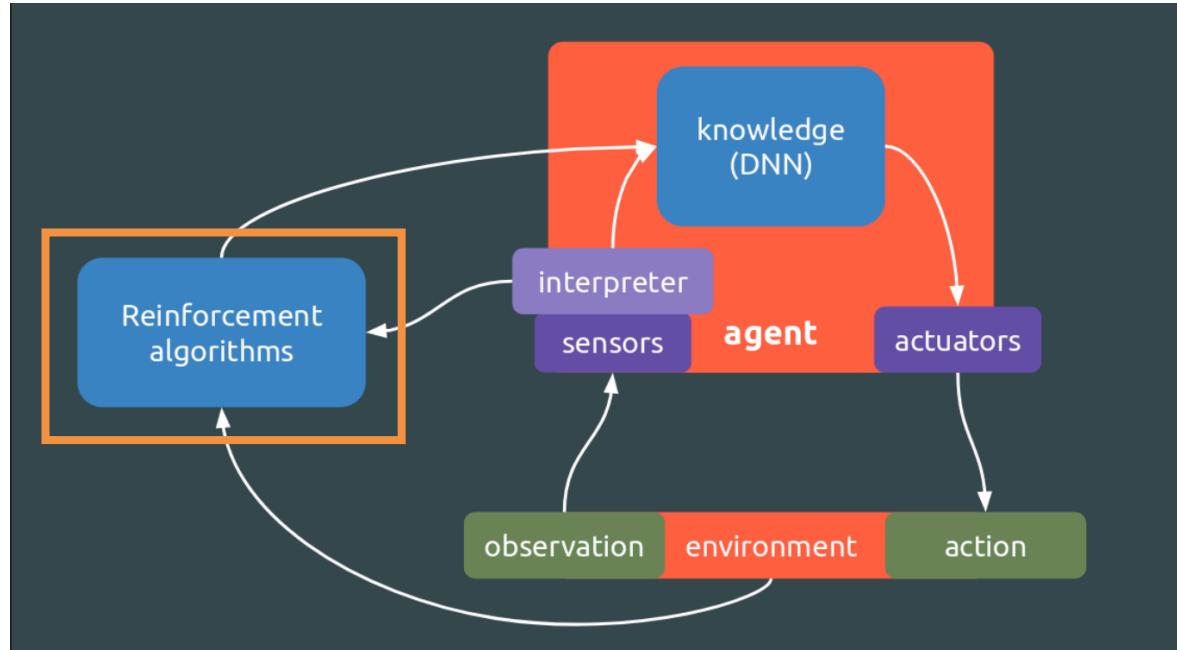
# Vista general



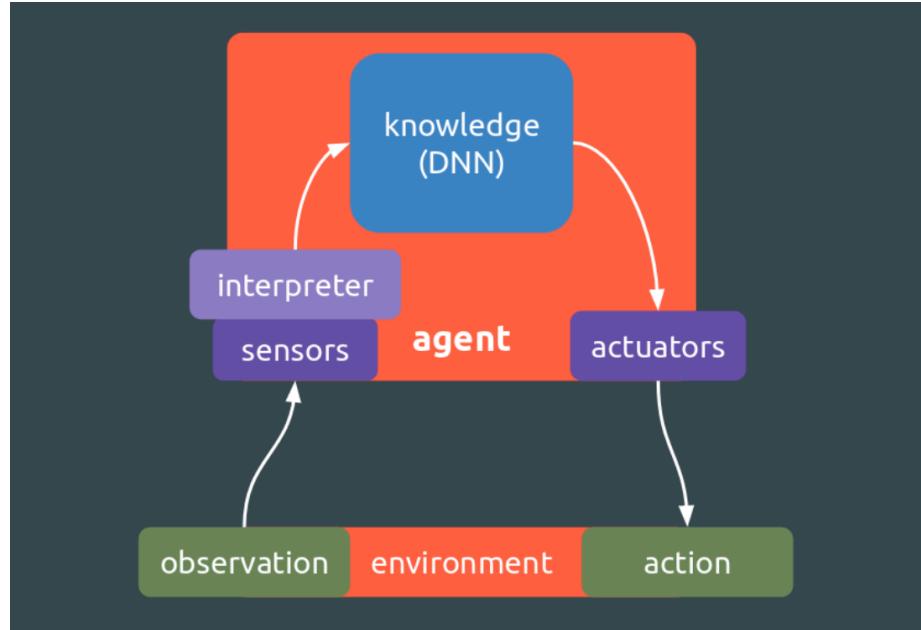
# Vista general



# Vista general



# Vista general



# Índice

---

¿Qué es aprendizaje por refuerzo?

Estado del arte

Distintos enfoques de aprendizaje por refuerzo

Retos actuales y de futuro

Vista general

**Conclusiones**

Bibliografía recomendada



## Conclusiones

---

- El aprendizaje por refuerzo es una de las ramas más prometedoras dentro de la inteligencia artificial.
- El proceso de aprendizaje se asemeja a la idea de aprendizaje que tenemos los seres humanos.
- Aunque es una rama muy prometedora, todavía están en sus primeros pasos. No está totalmente asentada y sus soluciones no se pueden aplicar en todos los campos y/o dominios.
- La unión de modelos de Deep Learning con algoritmos de aprendizaje por refuerzo ha sido la mezcla perfecta para abrir todo el conjunto de posibilidades con tanto potencial que tenemos a nuestro alcance actualmente.



# Índice

---

¿Qué es aprendizaje por refuerzo?

Estado del arte

Distintos enfoques de aprendizaje por refuerzo

Retos actuales y de futuro

Vista general

Conclusiones

**Bibliografía recomendada**



## Bibliografía recomendada

---

*“Reinforcement Learning: An introduction”, Sutton y Barto:*  
<http://incompleteideas.net/book/bookdraft2017nov5.pdf>

*Contenidos y documentación de OpenAI SpinningupRL:*  
<https://spinningup.openai.com/en/latest/>



# Gracias