

08MIAR-Aprendizaje por refuerzo

Sesión 1 – Introducción



Universidad
Internacional
de Valencia

De:



Planeta Formación y Universidades

Sobre mí

Julio Silva

Postdoctoral Researcher (<https://liviamtl.ca>)

juliojose.silva@professor.universidadviu.com

<https://es.linkedin.com/in/julio-silva-rodriguez>

Sobre la asignatura

El contenido de la asignatura se organizará en tres bloques:

- En el ***primer bloque*** introduciremos el **contexto** en el que se desarrollan las soluciones de aprendizaje por refuerzo, así como los **conceptos** y **términos** más importantes.
- El ***segundo bloque*** estará compuesto del estudio de los **algoritmos** que forman el estado del arte actual. Principalmente, cuáles son las estrategias de **aprendizaje**.
- El ***tercer bloque*** cubrirá una serie de sesiones **prácticas** donde implementaremos soluciones de aprendizaje por refuerzo usando diferentes **frameworks y librerías**.

Para más información sobre las sesiones: **Anexo de organización de sesiones en la sección de guía didáctica de la asignatura.**

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 sí que es recomendable.
- En la parte práctica trabajaremos con Python y algunas librerías típicas de entornos para trabajar con datos y modelado, como por ejemplo Keras, Tensorflow y Pytorch. Habrá disponible una sesión práctica grabada enfocada en la puesta a punto del entorno y la presentación de los frameworks que utilizaremos.
- En ambas partes será imprescindible conocer y tener cierta experiencia con modelos Deep Learning y, en concreto, con redes convolucionales.

Sobre la asignatura

El método de evaluación de la asignatura estará compuesto de:

- Participación en foros de debate (10%)
- Actividad sobre artículo científico (10%)
- Actividad sobre una serie de vídeos teóricos (10%)
- Proyecto práctico (30%)
- Examen final (40%)

Índice

¿Qué es el aprendizaje por refuerzo?

Estado del arte

Distintos enfoques de aprendizaje por refuerzo

Retos actuales y de futuro

Vista general

Conclusiones

Bibliografía recomendada

Índice

¿Qué es el aprendizaje por refuerzo?

Estado del arte

Distintos enfoques de aprendizaje por refuerzo

Retos actuales y de futuro

Vista general

Conclusiones

Bibliografía recomendada

¿Qué es el 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.”

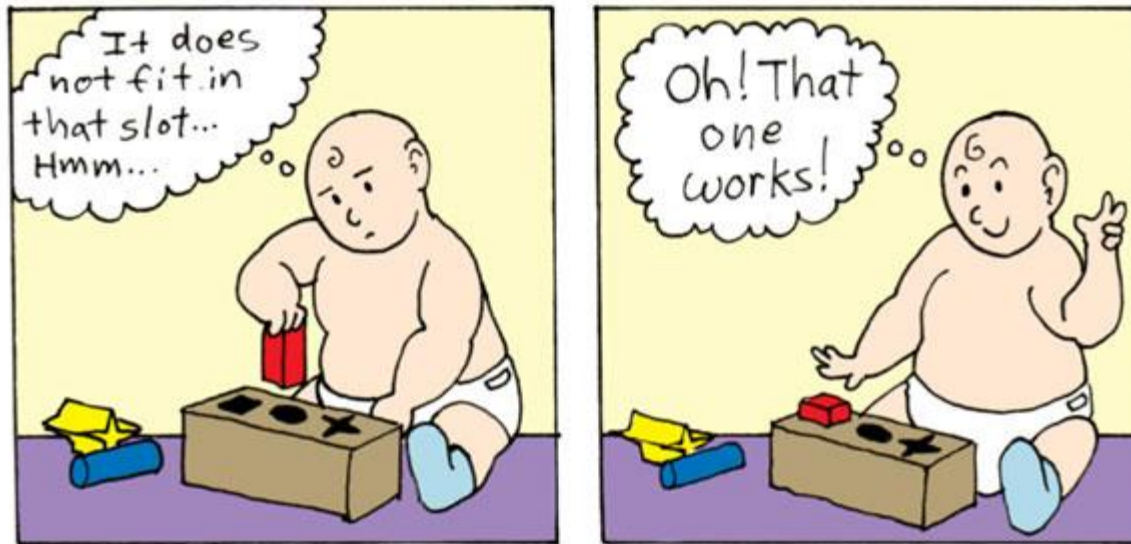
Reinforcement learning: An Introduction, R. Sutton & A. Barto

“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 el aprendizaje por refuerzo?

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



<https://rochemamabolo.files.wordpress.com/2018/07/trial02.jpg?w=4>

“Reinforcement learning is a framework for solving control tasks (also called decision problems) by building agents that learn from the environment by interacting with it through trial and error and receiving rewards (positive or negative) as unique feedback.”

<https://huggingface.co/learn/deep-rl-course/unit1/what-is-rl>

¿Qué es el 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 definir la 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 el 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 definir la 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 el 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 definir la 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 el 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.

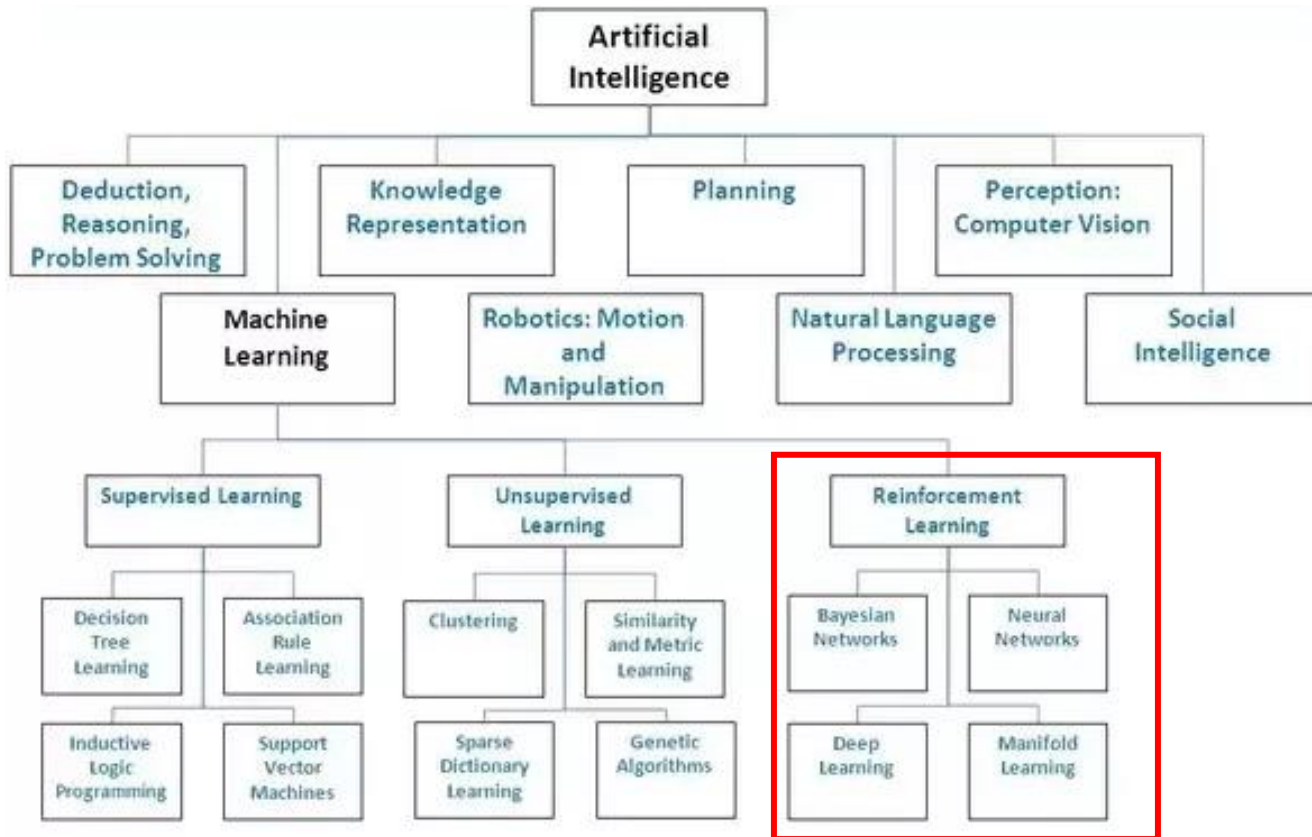
Desde un punto de vista de datos, podemos definir la relación de cada conjunto con el conocimiento que se puede obtener a partir de los datos:

Análisis descriptivo → Métodos no supervisados

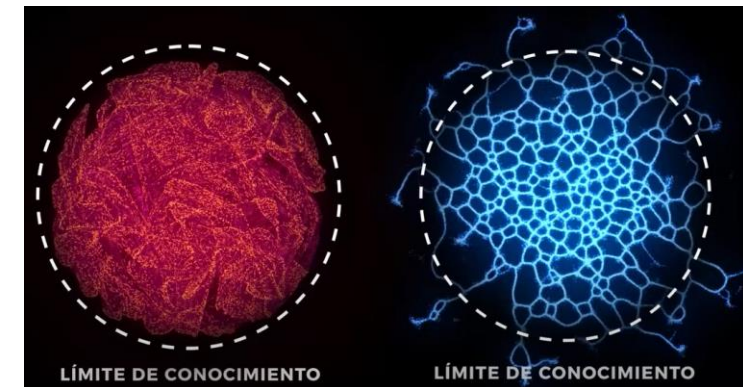
Análisis predictivo → Métodos supervisados

Análisis prescriptivo → Métodos aprendizaje por refuerzo

¿Qué es el aprendizaje por refuerzo?



<https://qph.fs.quoracdn.net/main-qimg-541146e2d9d611a42ebb5074aa72fef4.webp>



<https://www.youtube.com/watch?v=yAF3bopYMGk>

Índice

¿Qué es el 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 toda la información de la que disponen para aprender y adaptarse al problema que intentan resolver.

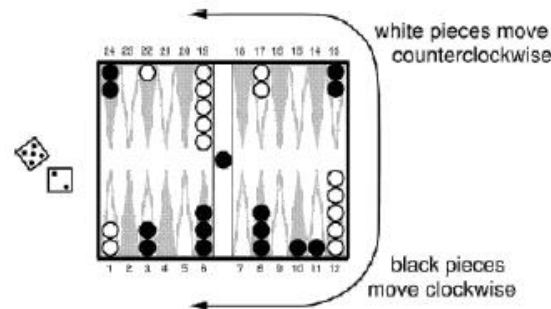
Estado del arte

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

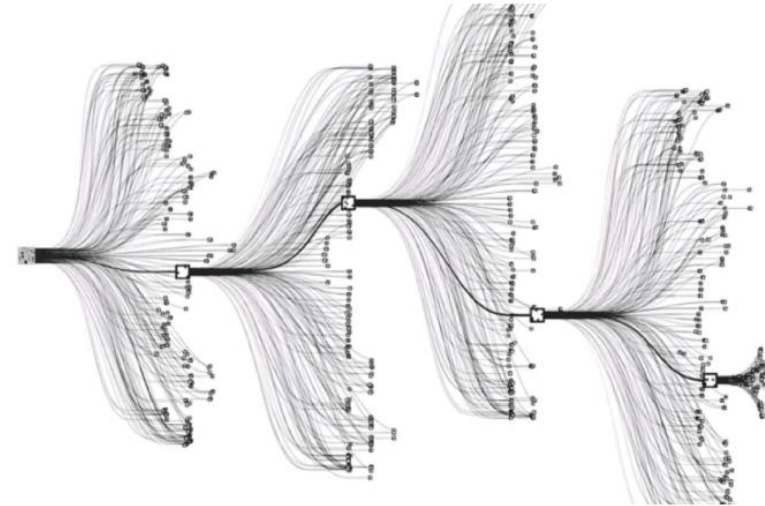
Vamos a centrarnos en Google Deepmind. El gran hito que consiguió Google Deepmind con su solución fue que combinó *Deep Learning* con Aprendizaje por refuerzo.

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.

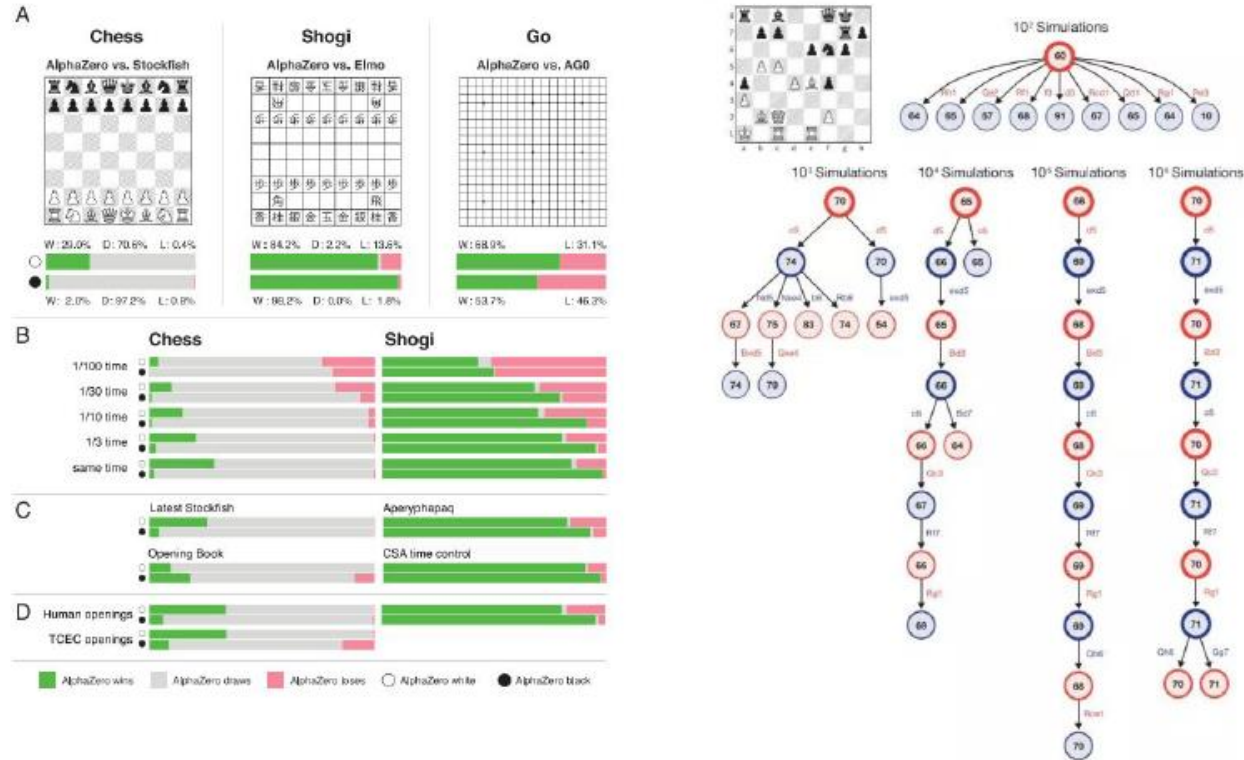


Estado del arte



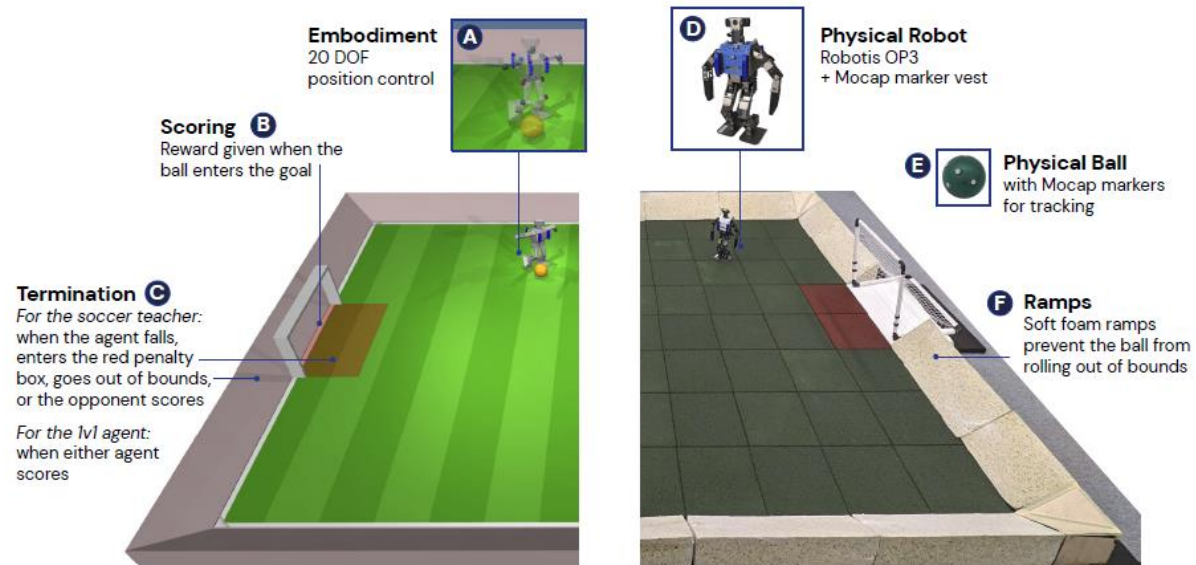
Hitos de DeepMind: **AlphaGo (2016)**

Estado del arte



Hitos de DeepMind: **AlphaZero (2017)**

Estado del arte



Hitos de DeepMind: **OP3-Soccer (2023)**

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*.

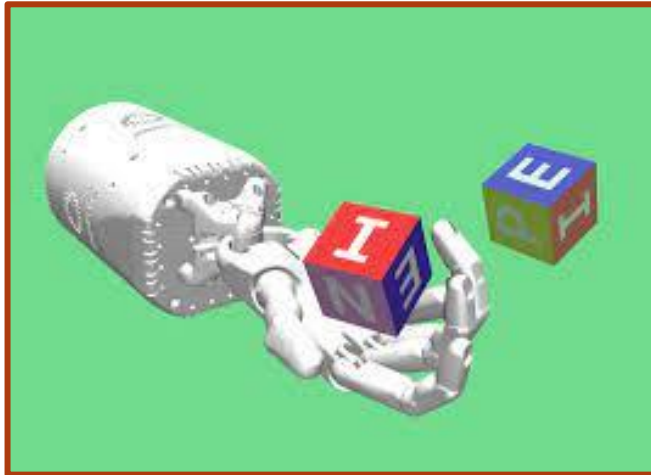


Estado del arte



Hitos de OpenAI: **Dota2 (2019)**

Estado del arte



Hitos de OpenAI: **Robótica (2019)**

Estado del arte: Tencencias Actuales



2023-4-27

Apr123: <https://arxiv.org/pdf/2304.13653.pdf>

Learning Agile Soccer Skills for a Bipedal Robot with Deep Reinforcement Learning

Tuomas Haarnoja^{*,1}, Ben Moran^{*,1}, Guy Lever^{*,1}, Sandy H. Huang^{*,1}, Dhruva Tirumala¹, Markus Wulfmeier¹, Jan Humplik¹, Saran Tunyasuvunakool¹, Noah Y. Siegel¹, Roland Hafner¹, Michael Bloesch¹, Kristian Hartikainen^{2,4}, Arunkumar Byravan¹, Leonard Hasenclever¹, Yuval Tassa¹, Fereshteh Sadeghi^{3,4}, Nathan Batchelor¹, Federico Casarini¹, Stefano Saliceti¹, Charles Game¹, Neil Sreendra, Kushal Patel, Marlon Gwira, Andrea Huber¹, Nicole Hurley¹, Francesco Nori¹, Raia Hadsell¹ and Nicolas Heess¹

^{*}Equal contributions, ¹DeepMind, ²University of Oxford, ³Google, ⁴Work done at DeepMind

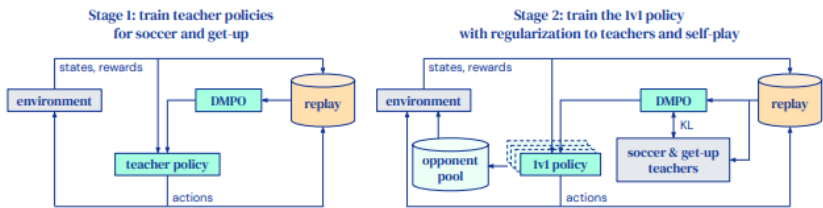
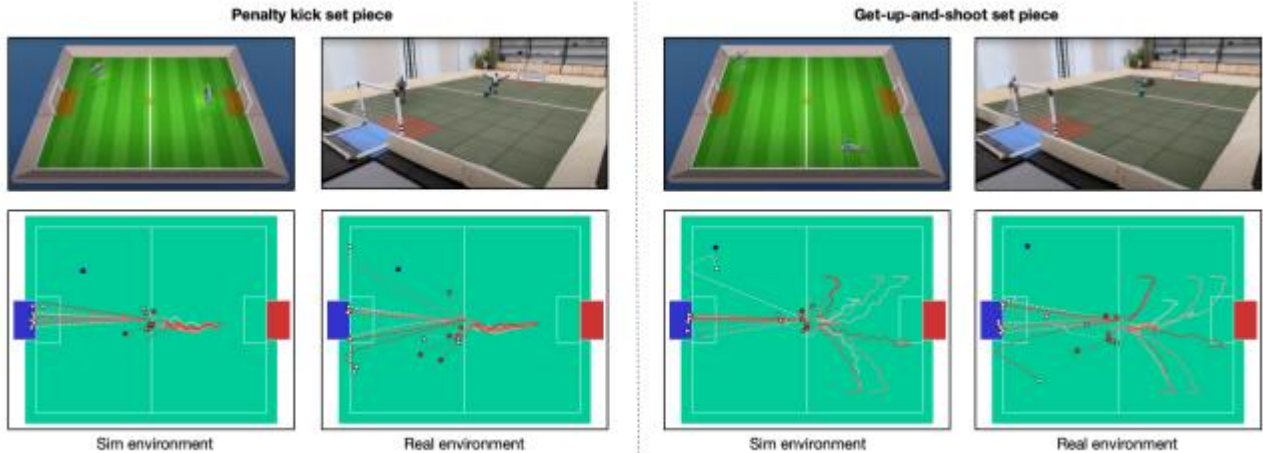


Figure 3 | We trained soccer agents in two stages. In the first stage (left), we train a separate soccer teacher and get-up teacher (Section 3.2.1). In the second stage (right), we distill these two teachers into a single agent that can both get up from the ground and play soccer (Section 3.2.2). The second stage also incorporates self-play: the opponent is uniformly randomly sampled from a pool that consists of policy snapshots from earlier in training. We found that this two-stage approach leads to qualitatively better behavior and improved sim-to-real transfer, compared to training an agent from scratch for the 1v1 soccer task.



Figure 4 | The joint key poses used to train the get-up teacher, extracted from a scripted get-up controller (Robotis, 2023).



Estado del arte: Tencencias Actuales

[nature](#) > [articles](#) > article

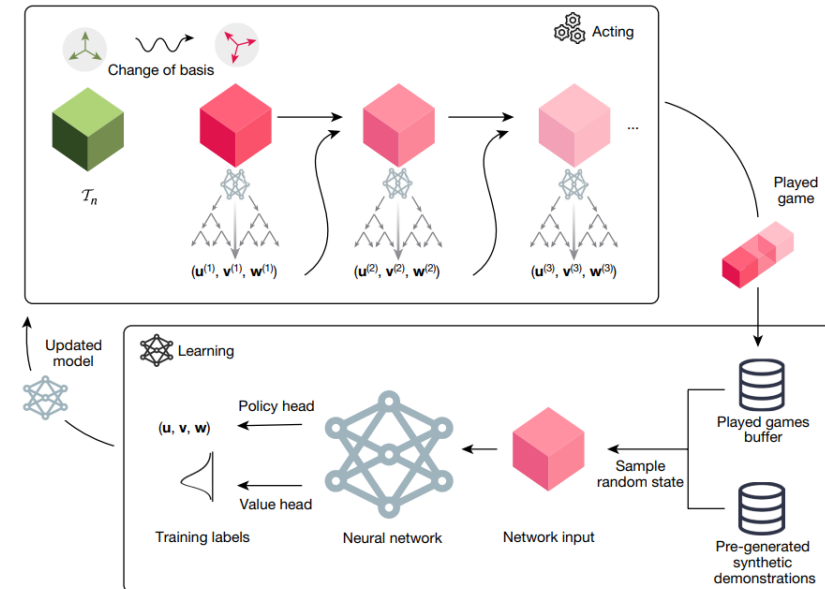
Article | [Open access](#) | [Published: 05 October 2022](#)

Discovering faster matrix multiplication algorithms with reinforcement learning

[Alhussein Fawzi](#) , [Matej Balog](#), [Aja Huang](#), [Thomas Hubert](#), [Bernardino Romera-Paredes](#), [Mohammadamin Barekatain](#), [Alexander Novikov](#), [Francisco J. R. Ruiz](#), [Julian Schrittwieser](#), [Grzegorz Swirszcz](#), [David Silver](#), [Demis Hassabis](#) & [Pushmeet Kohli](#)

[Nature](#) **610**, 47–53 (2022) | [Cite this article](#)

539k Accesses | **89** Citations | **3655** Altmetric | [Metrics](#)



Estado del arte: Tencencias Actuales

<https://huggingface.co/blog/rhlf>

Reinforcement Learning from Human Feedback (RLHF)

LE Can you explain to me how RL with human feedback works for language models like I am five?

AI Sure! So, imagine you have a robot named Rufus who wants to learn how to talk like a human. Rufus has a language model that helps him understand words and sentences.

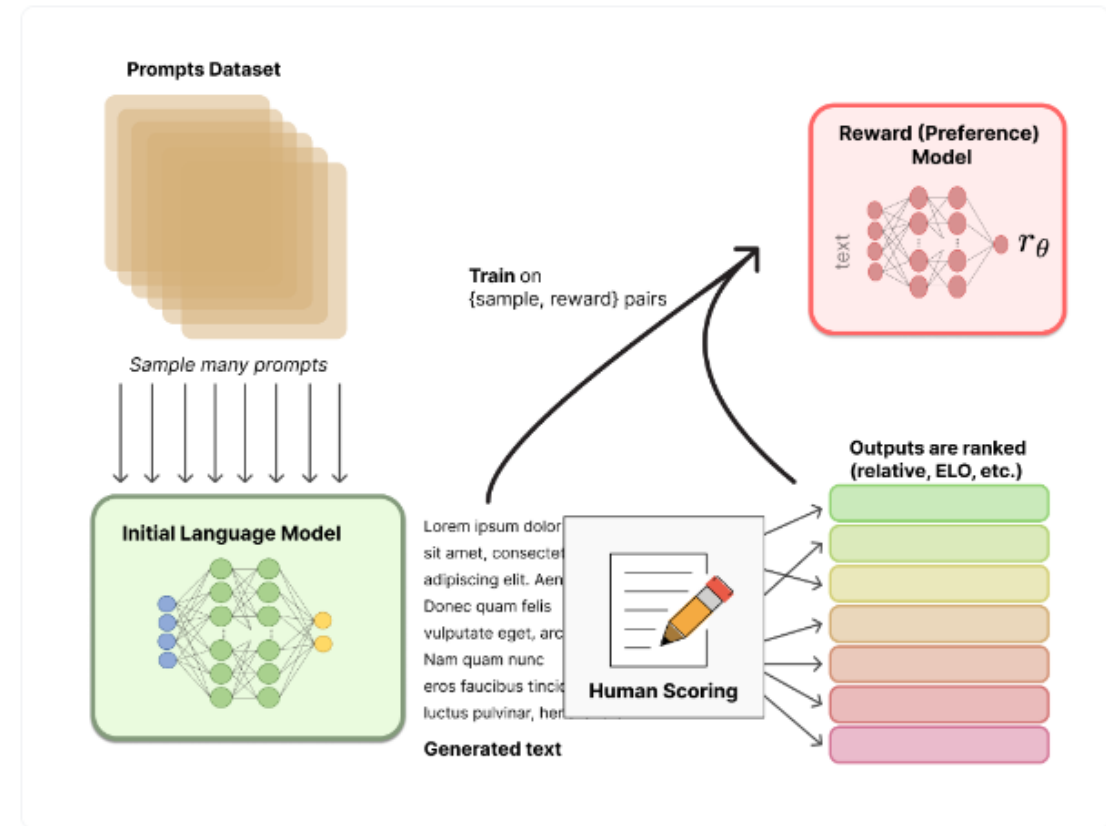
First, Rufus will say something using his language model. For example, he might say "I am a robot."

Then, a human will listen to what Rufus said and give him feedback on whether it sounded like a natural sentence a human would say. The human might say, "That's not quite right, Rufus. Humans don't usually say 'I am a robot.' They might say 'I'm a robot' or 'I am a machine.'"

Rufus will take this feedback and use it to update his language model. He will try to say the sentence again, using the new information he received from the human. This time, he might say "I'm a robot."

The human will listen again and give Rufus more feedback. This process will continue until Rufus can say sentences that sound natural to a human.

Over time, Rufus will learn how to talk like a human thanks to the feedback he receives from humans. This is how language models can be improved using RL with human feedback.



Estado del arte: Tencencias Actuales

Open X-Embodiment: Robotic Learning Datasets and RT-X Models

Open X-Embodiment Collaboration⁰

Abhishek Padalkar⁶, Acorn Pooley⁷, Ajinkya Jain¹⁰, Alex Bewley⁷, Alex Herzog⁷, Alex Irpan⁷, Alexander Khazatsky¹⁸, Anant Rai¹⁴, Anikait Singh^{7,22}, Anthony Brohan⁷, Antonin Raffin⁶, Ayzaan Wahid⁷, Ben Burgess-Limerick¹⁵, Beomjoon Kim¹², Bernhard Schölkopf¹³, Brian Ichter⁷, Cewu Lu^{17,9}, Charles Xu²², Chelsea Finn^{7,18}, Chenfeng Xu²², Cheng Chi^{8,18}, Chenguang Huang²⁴, Christine Chan⁷, Chuer Pan¹⁸, Chuyuan Fu⁷, Coline Devin⁷, Danny Driess⁷, Deepak Pathak², Dhruv Shah²², Dieter Buechler¹³, Dmitry Kalashnikov⁷, Dorsa Sadigh⁷, Edward Johns⁹, Federico Ceola¹¹, Fei Xia⁷, Freek Stulp⁶, Gaoyue Zhou¹⁴, Gaurav S. Sukhatme²⁰, Gautam Salhotra^{26,10}, Ge Yan²³, Giulio Schiavi⁴, Gregory Kahn²², Hao Su²³, Hao-Shu Fang¹⁷, Haochen Shi¹⁸, Heni Ben Amor¹, Henrik I. Christensen²³, Hiroki Furuta²¹, Homer Walke²², Hongjie Fang¹⁷, Igor Mordatch⁷, Ilija Radosavovic²², Isabel Leal⁷, Jacky Liang⁷, Jad Abou-Chakra¹⁵, Jaehyung Kim¹², Jan Peters¹⁹, Jan Schneider¹⁵, Jasmine Hsu⁷, Jeannette Bohg¹⁸, Jeffrey Bingham⁷, Jiajun Wu¹⁸, Jialin Wu⁸, Jianlan Luo²², Jiayuan Gu²⁹, Jie Tan⁷, Jihoon Oh²¹, Jitendra Malik²², Jonathan Tompson⁷, Jonathan Yang¹⁸, Joseph J. Lim¹², João Silvério⁶, Junhyek Han¹², Kanishka Rao⁷, Karl Pertsch^{22,18}, Karol Hausman⁷, Keegan Go¹⁰, Keerthana Gopalakrishnan⁷, Ken Goldberg²², Kendra Byrne⁷, Kenneth Oslund⁷, Kento Kawaharazuka²¹, Kevin Zhang², Krishan Rana¹⁵, Krishnan Srinivasan¹⁸, Lawrence Yunliang Chen²², Lerrel Pinto¹⁴, Liam Tan²², Lionel Ott⁴, Lisa Lee⁷, Masayoshi Tomizuka²², Maximilian Du¹⁸, Michael Ahn⁷, Mingtong Zhang²⁵, Mingyu Ding²², Mohan Kumar Srirama², Mohit Sharma², Moo Jin Kim¹⁸, Naoaki Kanazawa²¹, Nicklas Hansen²³, Nicolas Heess⁷, Nikhil J. Joshi⁷, Niko Suenderhauf¹⁵, Norman Di Palo⁹, Nur Muhammad Mahi Shafiullah¹⁴, Oier Mees²⁴, Oliver Kroemer⁷, Pannag R. Sanketi⁷, Paul Wohlhart⁷, Peng Xu⁷, Pierre Sermanet⁷, Priya Sundareshan¹⁸, Quan Vuong⁷, Rafael Rafailov^{7,18}, Ran Tian²², Ria Doshi²², Roberto Martin-Martin²⁰, Russell Mendonca², Rutav Shah²⁰, Ryan Hoque²², Ryan Julian⁷, Samuel Bustamante⁹, Sean Kirmani⁷, Sergey Levine^{7,22}, Sherry Moore⁷, Shikhar Bahl⁷, Shivin Dass^{26,20}, Shubham Sonawani¹, Shuran Song⁸, Sichun Xu⁷, Siddhant Halder¹⁴, Simeon Adebola²², Simon Guist¹³, Sororous Nasiriany²⁰, Stefan Schaal¹⁰, Stefan Welker⁷, Stephen Tian¹⁸, Sudeep Dasari², Suneel Belkale¹⁸, Takayuki Osa²¹, Tatsuya Harada^{21,16}, Tatsuya Matsushima²¹, Ted Xiao⁷, Tianhe Yu⁷, Tianli Ding⁷, Todor Pavchev⁷, Tony Z. Zhao¹⁸, Travis Armstrong⁷, Trevor Darrell²², Vidhi Jain^{7,2}, Vincent Vanhoucke⁷, Wei Zhan²², Wenxuan Zhou^{7,2}, Wolfram Burgard²⁷, Xi Chen⁷, Xiaolong Wang²³, Xinghao Zhu²², Xuanlin Li²³, Yao Lu⁷, Yevgen Chebotar⁷, Yifan Zhou¹, Yifeng Zhu²⁰, Ying Xu⁷, Yixuan Wang²⁵, Yonatan Bisk², Yoonyoung Cho¹², Youngwoon Lee²², Yuchen Cui¹⁸, Yueh-Hua Wu²³, Yujin Tang^{7,21}, Yuke Zhu²⁰, Yunzhu Li²⁰, Yusuke Iwasawa²¹, Yutaka Matsuo²¹, Zhuo Xu⁷, Zichen Jeff Cui¹⁴

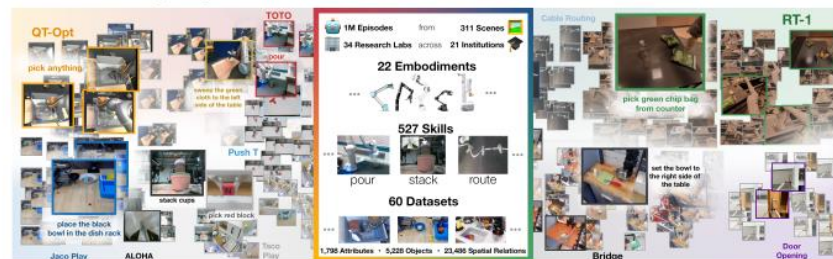


Fig. 1: We propose an open, large-scale dataset for robot learning curated from 21 institutions across the globe. The dataset represents diverse behaviors, robot embodiments and environments, and enables learning generalized robotic policies.

VOYAGER: An Open-Ended Embodied Agent with Large Language Models

Guanzhi Wang^{1,2}, Yuqi Xie³, Yunfan Jiang^{1*}, Ajay Mandlekar^{1*}, Chaowei Xiao^{1,5}, Yuke Zhu^{1,3}, Linxi "Jim" Fan^{1,4}, Anima Anandkumar^{1,2†}

¹NVIDIA, ²Caltech, ³UT Austin, ⁴Stanford, ⁵UW Madison

*Equal contribution [†]Equal advising [‡]Corresponding authors

<https://voyager.minedojo.org>

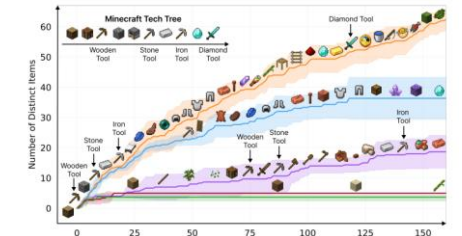
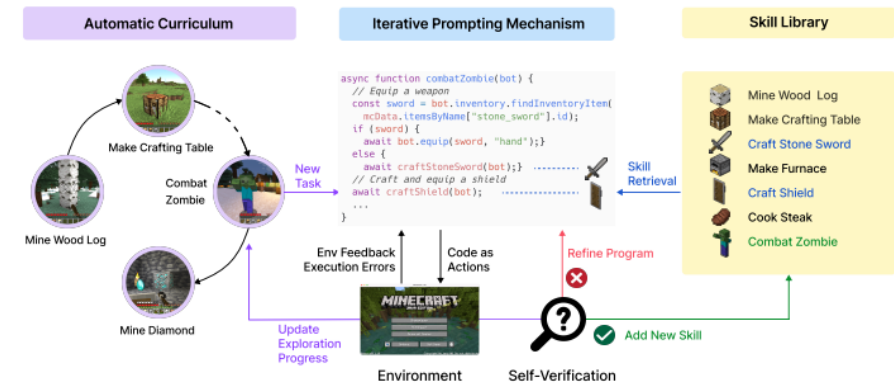


Figure 1: VOYAGER discovers new Minecraft items and skills continually by self-driven exploration, significantly outperforming the baselines. X-axis denotes the number of prompting iterations.



May23: <https://arxiv.org/pdf/2305.16291.pdf>

Oct23: <https://robotics-transformer-x.github.io/paper.pdf>

Estado del arte: Tencencias Actuales

SayTap: Language to Quadrupedal Locomotion

Yujin Tang

yujintang@google.com
Google DeepMind

Wenhao Yu

magicmelon@google.com
Google DeepMind

Jie Tan

jietan@google.com
Google DeepMind

Heiga Zen

heigazen@google.com
Google DeepMind

Aleksandra Faust

sandrafaust@google.com
Google DeepMind

Tatsuya Harada

harada@mi.t.u-tokyo.ac.jp
The University of Tokyo

Abstract: Large language models (LLMs) have demonstrated the potential to perform high-level planning. Yet, it remains a challenge for LLMs to comprehend low-level commands, such as joint angle targets or motor torques. This paper proposes an approach to use foot contact patterns as an interface that bridges human commands in natural language and a locomotion controller that outputs these low-level commands. This results in an interactive system for quadrupedal robots that allows the users to craft diverse locomotion behaviors flexibly. We contribute an LLM prompt design, a reward function, and a method to expose the controller to the feasible distribution of contact patterns. The results are a controller capable of achieving diverse locomotion patterns that can be transferred to real robot hardware. Compared with other design choices, the proposed approach enjoys more than 50% success rate in predicting the correct contact patterns and can solve 10 more tasks out of a total of 30 tasks. (<https://saytap.github.io>)

Keywords: Large language model (LLM), Quadrupedal robots, Locomotion



Jun23: <https://arxiv.org/abs/2306.07580>

Índice

¿Qué es el 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 dos grupos principales:

Entorno puramente digital - Ámbito de investigación

Entorno Digital-a-Real - Ámbito de investigación/aplicación

Vamos a entrar en detalle con cada enfoque.

Distintos enfoques del aprendizaje por refuerzo



Entorno puramente digital - Ámbito de Investigación

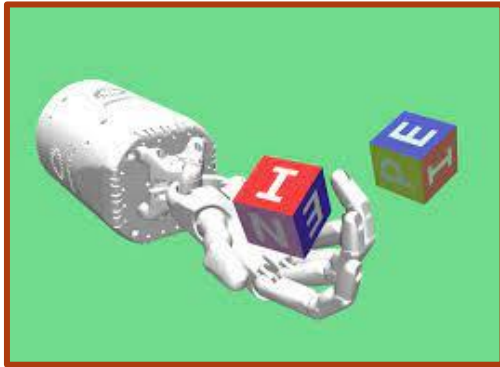
Este es el entorno típico que nos encontraremos en la mayoría de ejemplos a día de hoy.

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.

Índice

¿Qué es el 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

A pesar de tener en la actualidad 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

Aun 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 los 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 el aprendizaje por refuerzo?

Estado del arte

Distintos enfoques de aprendizaje por refuerzo

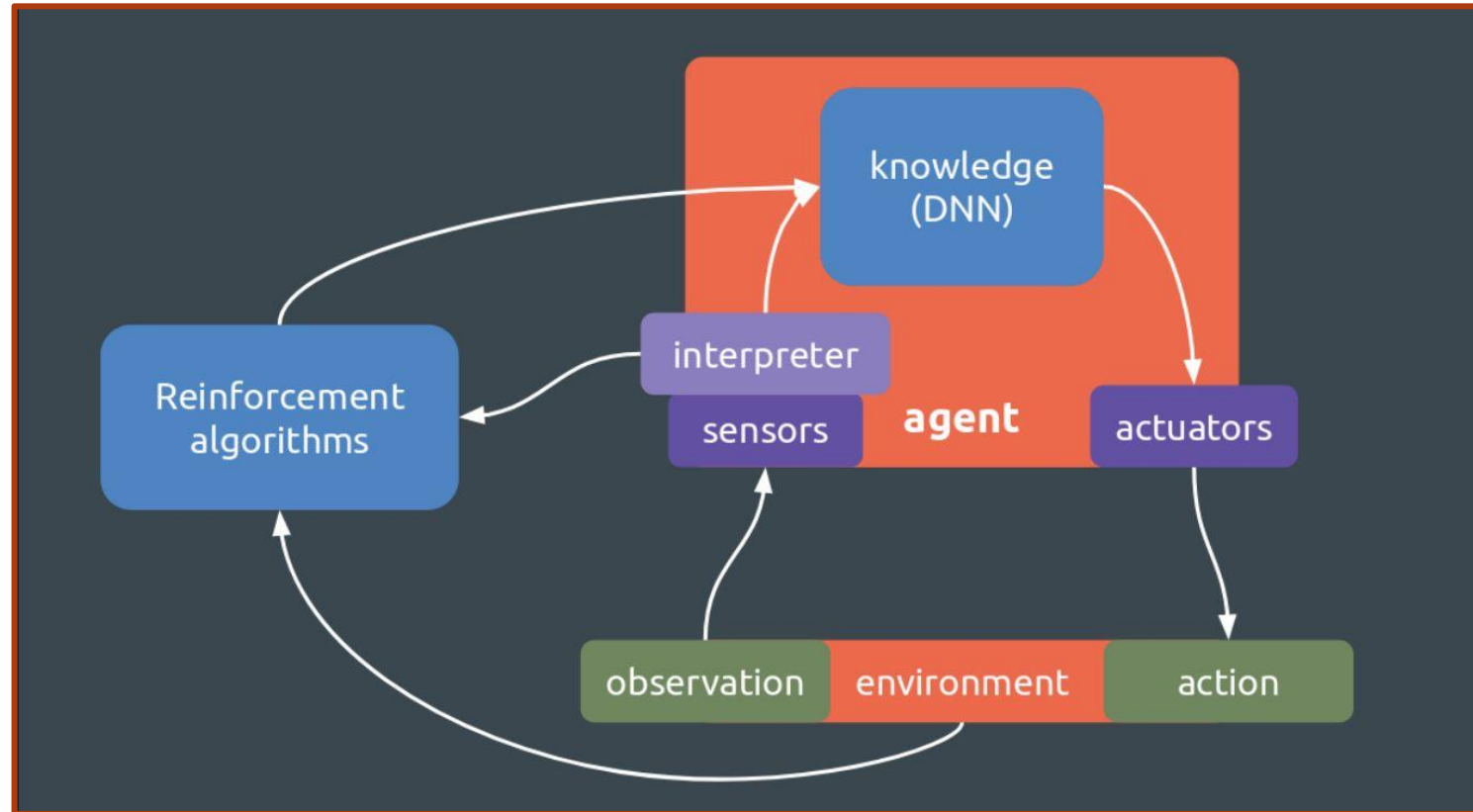
Retos actuales y de futuro

Vista general

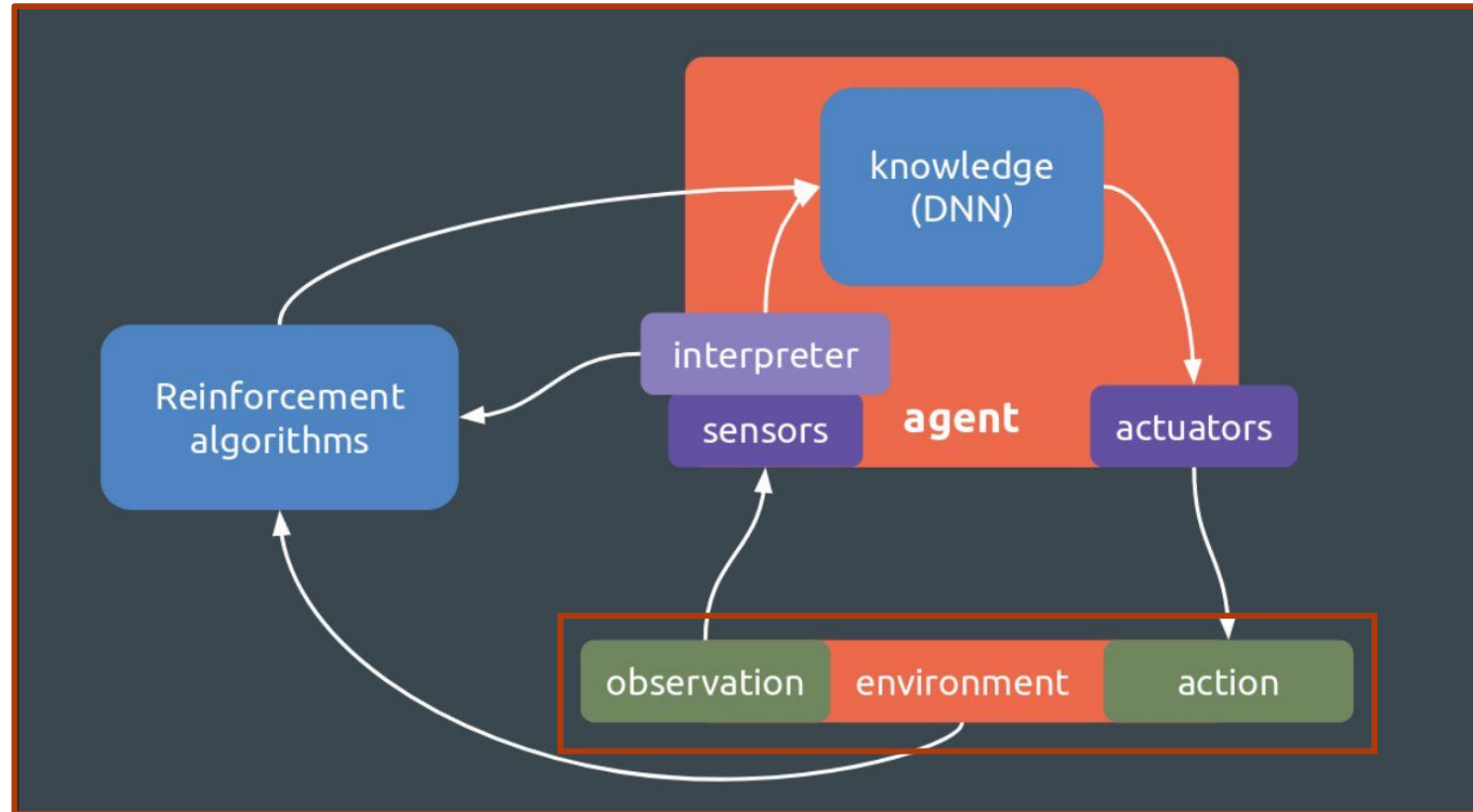
Conclusiones

Bibliografía recomendada

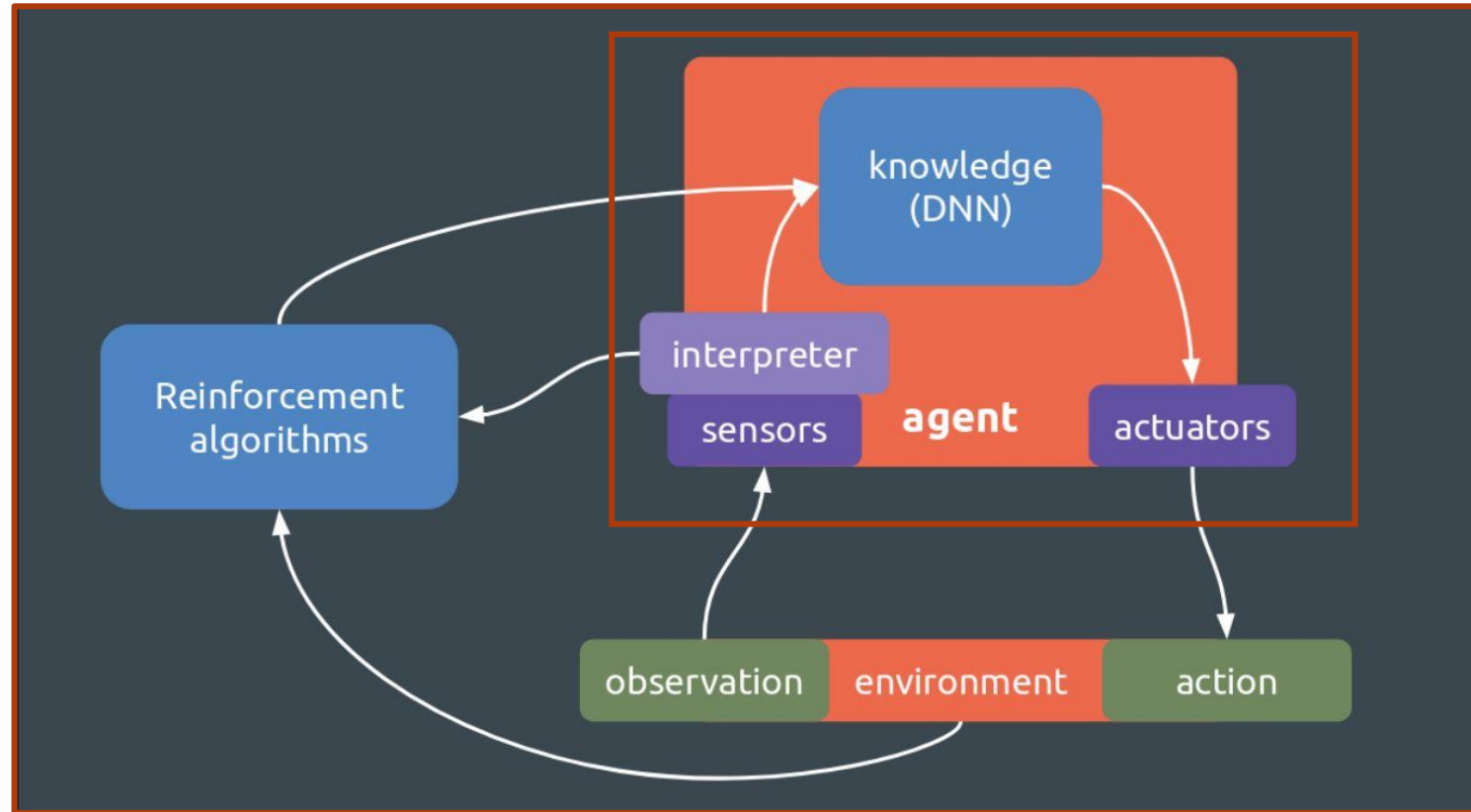
Vista general



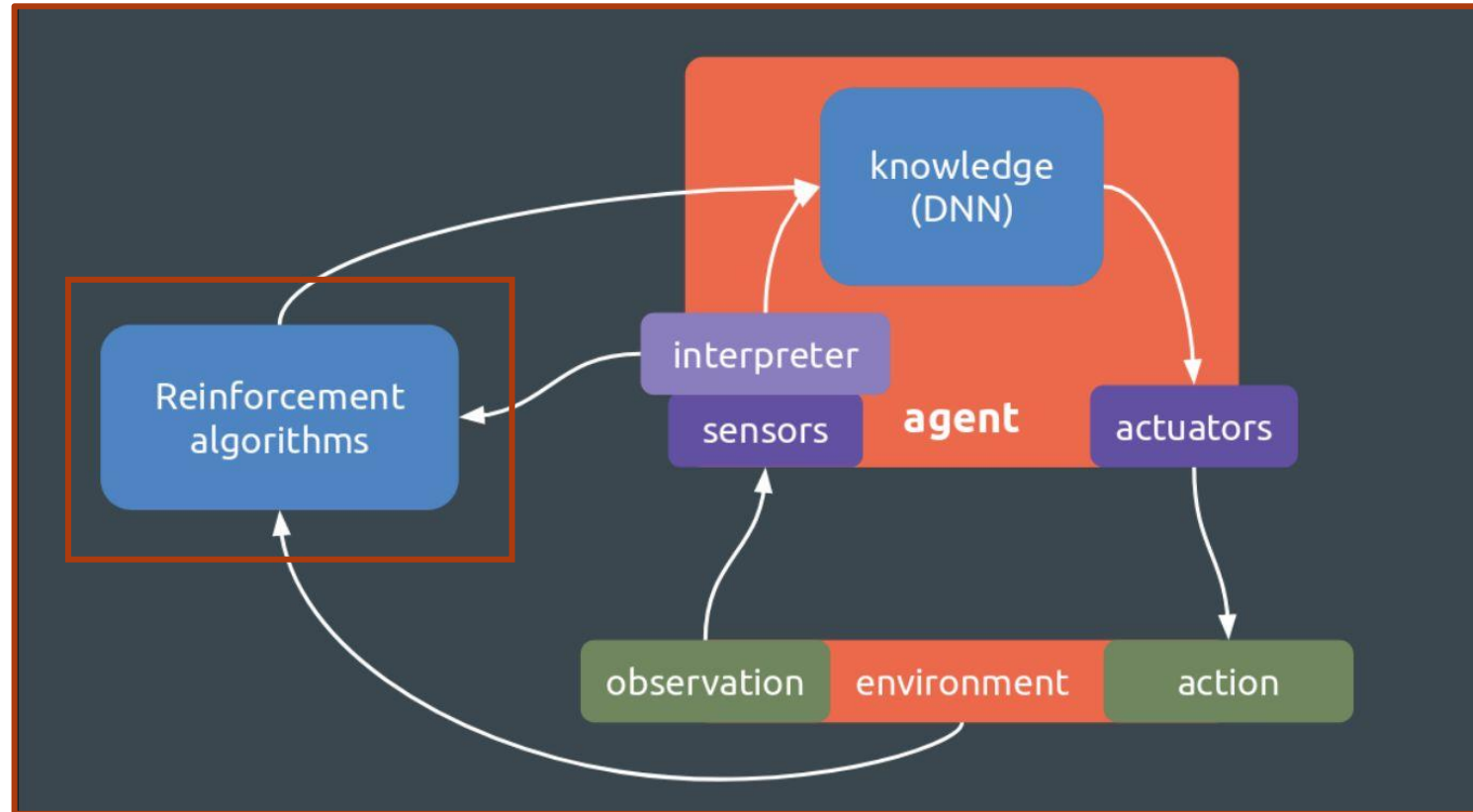
Vista general



Vista general



Vista general



Índice

¿Qué es el 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 el 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/>

HuggingFace Courses:
<https://huggingface.co/learn/deep-rl-course/>



viu

Universidad
Internacional
de Valencia

universidadviu.com

De:
 Planeta Formación y Universidades