

UNIVERSIDAD CATÓLICA SANTO TORIBIO DE MOGROVEJO
FACULTAD DE INGENIERÍA
ESCUELA DE INGENIERÍA DE SISTEMAS Y COMPUTACIÓN



**Aplicaciones de algoritmos de aprendizaje reforzado profundo: una
revisión sistemática de la literatura**

**TRABAJO DE INVESTIGACIÓN PARA OPTAR EL GRADO ACADÉMICO DE
BACHILLER EN INGENIERÍA DE SISTEMAS Y COMPUTACIÓN**

AUTOR

Sara Maria Benel Ramirez

ASESOR

Karla Cecilia Reyes Burgos

<https://orcid.org/0000-0003-3520-5076>

Chiclayo, 2022

Índice

Resumen	3
Abstract	4
Introducción.....	5
Metodología.....	6
Criterios de inclusión y exclusión	7
Evaluación de la calidad	7
Recopilación de datos.....	8
Resultados y discusión	12
Conclusiones	13
Agradecimientos	13
Referencias	14

Resumen

El objetivo de este estudio es realizar un análisis sistemático de las aplicaciones que se da a los algoritmos de aprendizaje reforzado profundo para identificar cuales están siendo utilizados actualmente y para qué. Con el fin de alcanzar este propósito, se ha realizado una revisión sistemática de los artículos de investigación que resultaron de la búsqueda, y el filtrado, en las bases de datos ScienceDirect, IEEE Xplore y ProQuest. Inicialmente se obtuvieron 926 artículos de los cuales solo 8 pasaron todos los filtros establecidos. Después de su lectura se logró identificar cinco algoritmos los cuales fueron empleados en los sectores de redes, planeamiento de rutas, programación de trabajos y administración de carga. Finalmente, en esta revisión, estos algoritmos demuestran ser de mucha utilidad y tener la capacidad de brindar solución a diversos problemas; y por esta razón se espera que se sigan investigando formas de aplicar estos algoritmos.

Palabras clave: aplicación, aprendizaje reforzado profundo, aprendizaje reforzado.

Abstract

The objective of this study is to carry out a systematic analysis of the applications given to deep reinforced learning algorithms to identify which are currently being used and what for. In order to achieve this purpose, a systematic review of the research articles that resulted from the search, and the filtering, has been carried out in the ScienceDirect, IEEE Xplore and ProQuest databases. Initially, 926 articles were obtained, of which only 8 passed all the established filters. After reading it, it was possible to identify five algorithms which were used in the networks, route planning, job scheduling and load management sectors. Finally, in this review, these algorithms prove to be very useful and have the ability to provide solutions to various problems; For this reason, it is expected that more ways of applying these algorithms will be investigated.

Keywords: application, deep reinforcement learning, reinforcement learning.

Introducción

El aprendizaje automático (Machine Learning, ML) es una rama de la inteligencia artificial (Artificial Intelligence, AI) que ha tomado popularidad estos últimos años, este campo se caracteriza porque su objetivo principal es hacer que las computadoras aprendan. En el proceso de aprendizaje se desarrolla una etapa de entrenamiento en la que un agente – la máquina – puede adquirir conocimiento y aplicarlo. Este entrenamiento se puede dar de distintas formas, pero en esta rama ya se destacan tres metodologías de enseñanza que han dado resultados prometedores: el aprendizaje supervisado, no supervisado y reforzado.

El aprendizaje reforzado (Reinforcement Learning, RL) destaca su utilidad por la forma peculiar en que adiestra a sus agentes inteligentes. Este método de instrucción no tiene la necesidad, a diferencia de las otras dos, de requerir grandes cantidades de datos para realizar su entrenamiento. Su forma de entrenar consiste en aprender a través de la prueba y el error [1].

Esta metodología del aprendizaje ha concebido una gran innovación con la fusión que ha tenido con las redes neuronales (Neural Networks, NN), específicamente con las redes neuronales profundas (Deep Neural Networks, DNN). Una red neuronal es básicamente un conjunto de capas de perceptrones llamados “neuronas”; el término red neural profunda hace entonces referencia a una gran cantidad de estas capas [2]. El resultado de la utilización de DNNs en RL ha dado paso al aprendizaje reforzado profundo (Deep Reinforcement Learning, DRL) y este a su vez a grandes avances en el desarrollo de algoritmos de DRL.

Los algoritmos de DRL han dado que hablar en la comunidad del ML debido a su gran capacidad para volverse expertos en aquello que aprenden [3] [4], y aunque su desarrollo va en aumento aún no se ha podido observar su aplicación para resolver problemas reales. Es por esto que el objetivo de este documento es el de hacer una revisión sistemática de la literatura sobre las aplicaciones que se les dan a estos algoritmos. De esta manera se pretende identificar algoritmos de DRL y reconocer sus usos actuales.

Metodología

En este apartado se plantean las preguntas que se pretenden responder en esta revisión, se muestra el diseño de la estrategia de búsqueda de manera detallada y el seguimiento de filtrado que nos dará como resultado los artículos a analizar. Para esto se utilizó la propuesta de Kitchenham [5] quien plantea seguir los siguientes acápites:

Preguntas de investigación

Las preguntas se plantearon teniendo en cuenta el objetivo de esta revisión.

- RQ1. ¿Qué algoritmos de DRL fueron usados en estas aplicaciones?
- RQ2. ¿Qué aplicaciones les dan a estos algoritmos?

Diseño de la estrategia de búsqueda

En esta sección se determinan las bases de datos que se usaron, la cadena de búsqueda y el proceso de búsqueda.

Bases de datos

Se han consultado 3 bases de datos de investigaciones científicas: Science Direct, IEEE Xplore, ProQuest.

Términos de búsqueda

La cadena de búsqueda usada para la consulta en las bases de datos se estructuró mediante el siguiente proceso:

- a) Se eligieron sustantivos representativos y que engloben el tema:
 - Deep reinforcement learning
 - Use
 - Application
- b) Estas palabras se unieron de la manera adecuada para indicar la relación entre estos enunciados y se llegó a formar dos cadenas de búsqueda que se usarán en distintas bases de datos. Se colocaron entre comillas porque de esta manera se aplicarán en las bases de datos:
 - “Using deep reinforcement learning”
 - “Application of deep reinforcement learning”

Proceso de búsqueda

La búsqueda inicial a través de las bases de datos se realizó el 13 de julio del 2020 empleando diferentes cadenas para diferentes bases de datos.

Tabla 1: Cadenas de búsqueda en las BBDD

Cadena de búsqueda	Base de datos
“Using deep reinforcement learning”	ProQuest, IEEE Xplore
“Application of Deep reinforcement learning”	ScienceDirect

Esta búsqueda nos da un total de 926 resultados y se puede observar en la tabla 2. Esta cadena se colocó entre comillas para una búsqueda más exacta.

Tabla 2: Resultados de la búsqueda inicial en las BBDD

ScienceDirect	IEEE Xplore	ProQuest
19	231	676

Criterios de inclusión y exclusión

Estos criterios se usarán para establecer un marco de búsqueda y filtrar los datos obtenidos.

Criterio de inclusión

- Solo se seleccionarán artículos de investigación
- Los artículos deberán estar en el idioma inglés o español
- Solo se aceptarán artículos en el rango del 2016 al 2020
- Solo se aceptarán artículos que contengan las palabras de la cadena de búsqueda en el título, abstract o palabras clave del autor.

Criterio de exclusión

- Publicaciones duplicadas
- No pertenecen a revistas científicas
- Se excluirán si no son artículos de acceso libre

Evaluación de la calidad

Con el fin de salvaguardar la calidad de la información se establecieron criterios.

- No se seleccionarán artículos de menos de 5 páginas
- Solo se seleccionarán artículos evaluados por expertos, si la herramienta de búsqueda nos facilita esta opción.
- No se seleccionarán artículos que no definan claramente su objetivo.
- No se seleccionarán artículos en dónde se proponga un nuevo algoritmo.
- No se seleccionarán artículos en donde no quede claro a que rama pertenece su aplicación o no lo usen.
- No se seleccionarán artículos que no especifiquen qué algoritmo de aprendizaje reforzado profundo usaron para su aplicación.

Recopilación de datos

En el curso de la obtención de los artículos de revisión se procedió a realizar una búsqueda avanzada con los criterios de inclusión y exclusión. Cada buscador en las distintas bases de datos nos ofrecía diferentes parámetros. A continuación, se detallará la búsqueda avanzada realizada en cada uno de ellos:

ScienceDirect

En esta base de datos después de la búsqueda inicial se procedió a filtrar por el rango de años y no se evidenció un cambio en el número de resultados iniciales, lo cual indica que todos fueron publicados entre los años 2016 y 2020. El tercer filtro nos muestra una reducción de 12 obteniendo así 7 resultados. Después del cuarto filtro no se observan cambios, mientras que después de filtrar por artículos de acceso libre obtenemos, finalmente, 1 artículo apto para evaluar. Este proceso se observa mejor en la Figura 1.

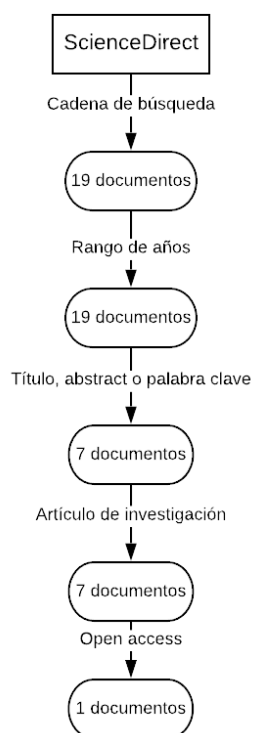


Figura 1: Proceso de filtrado en ScienceDirect

IEEE Xplore

El filtrado de la búsqueda avanzada se inició buscando “using deep reinforcement learning” en el abstract, del cual se obtuvo 104 documentos. De estos últimos, 20 eran journals y finalmente se identificó que solo 10 eran de acceso libre.

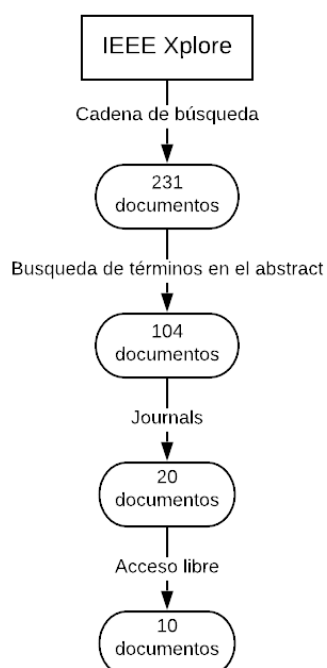


Figura 2: Proceso de filtrado en IEEE Xplore

ProQuest

La búsqueda avanzada en esta base de datos se realizó de manera parecida a la empleada en IEEE Xplore. Después de obtener 676 resultados se procedió a buscar la cadena de búsqueda en el abstract, sin omitir las comillas, y se obtuvieron 181 resultados de los cuales 14 eran de revistas científicas. Posterior a aplicar el filtro por texto completo, artículos evaluados por expertos y artículos se obtuvo 1 documento.

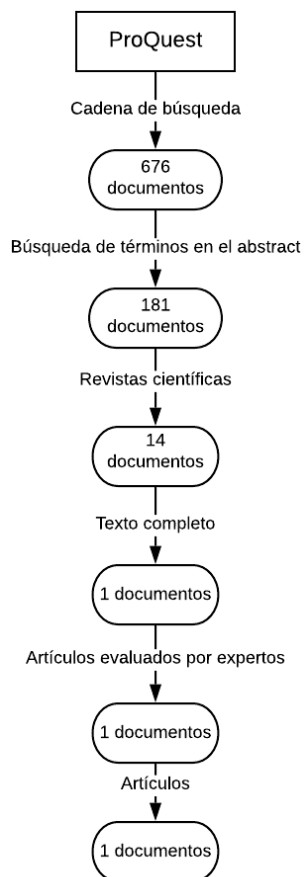


Figura 3: Proceso de filtrado en ProQuest

Después de aplicar la búsqueda avanzada se obtuvo un total de 12 artículos. A estos se le aplicó la evaluación de calidad y el proceso fue el que se puede apreciar en la figura 4. Este filtrado se realizó de manera grupal para todos los artículos encontrados.

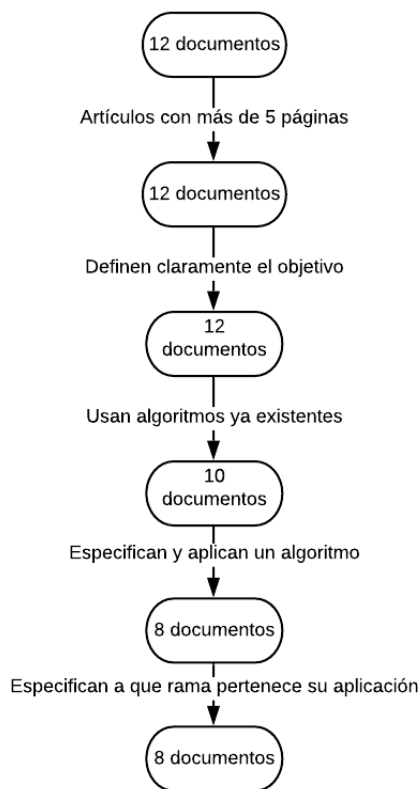


Figura 4: Proceso de evaluación de calidad

Concluyendo con esta etapa se adquirieron 8 artículos que nos ayudarán a responder las preguntas planteadas. El proceso completo de filtrado se puede observar en la figura 5 de manera resumida.

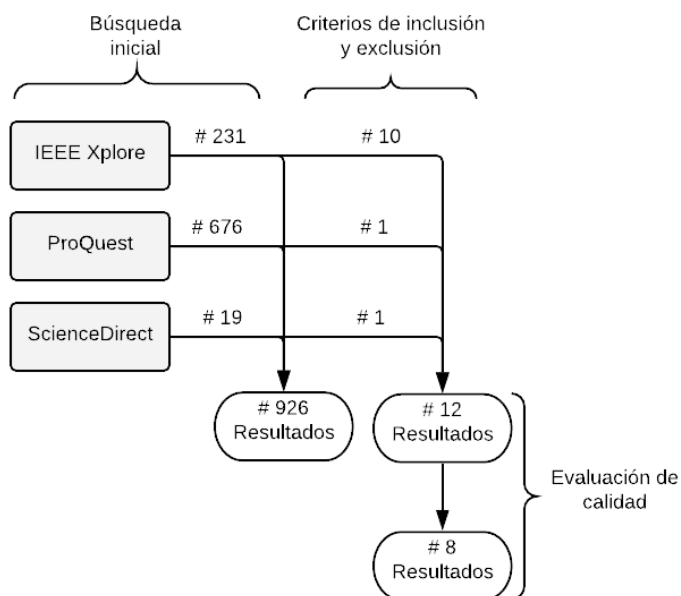


Figura 5: Proceso de búsqueda y filtrado

Resultados y discusión

Después de la lectura y la recolección de información pertinente para dar solución a las preguntas planteadas se procederá a responderlas en esta sección con el uso de los artículos seleccionados que podemos observar en la tabla 3.

Tabla 3: Artículos seleccionados

BBDD	Artículos seleccionados
IEEE Xplore	[6] [7]–[12]
ProQuest	[13]

Algoritmos de DRL utilizados (RQ1)

Como se puede observar en la tabla 4 se identificaron cinco tipos de algoritmos usados en los artículos. Es necesario mencionar que estos algoritmos, en su mayoría, para poder aplicarse a cierta problemática se tuvieron que modificar. Sin embargo, estas modificaciones no logran cambiar la naturaleza del algoritmo y por el contrario, es un cambio necesario para que pueda aplicarse en distintos sectores y a diversas problemáticas.

Tabla 4: Artículos y algoritmos

Artículo	Algoritmo usado
[10] [6] [8] [13]	Deep Q-Network (DQN)
[9]	Ape-X
[11]	Deep Deterministic Policy Gradient (DDPG)
[12]	Deep Q-learning
[7]	Q-learning

Después de organizar los resultados se observa que el algoritmo más utilizado fue el algoritmo desarrollado por la empresa DeepMind, de Google, llamado DQN [4].

Aplicaciones de algoritmos de DRL (RQ2)

Para responder a la segunda interrogante se identificaron las aplicaciones que se le dieron a los algoritmos anteriormente mencionados y se organizaron los resultados por artículo en la tabla 5. En ella podemos observar que solo en dos artículos se busca la misma aplicación y las demás están estrechamente relacionadas.

Tabla 5: Aplicaciones y artículos

Aplicación	Artículos
Path planning	[7], [11]
Mobile edge computing	[6]
Radio networks	[10]
Virtual network	[8]
Load management	[13]
Job scheduling	[12]
Network slicing	[9]

Para una mejor organización, visualización y comprensión de las aplicaciones se las agrupó en sectores más generales, estas clasificaciones fueron determinadas por el autor y se las puede observar en la tabla 6 con sus respectivos artículos. En esta tabla se puede evidenciar que en donde más se aplican los algoritmos de DRL es en las redes.

Tabla 6: Aplicaciones en sectores generales

Clasificación de aplicaciones	Artículos
Path planning	[7], [11]
Networks	[6], [8]–[10]
Load management	[13]
Job scheduling	[12]

Conclusiones

En esta revisión se han analizado 8 artículos de investigación para dar respuesta a las interrogantes inicialmente planteadas. Se ha mencionado a los algoritmos que se usaron en estas investigaciones, específicamente: DQN, DDPG, Ape-X, Deep Q-learning y Q-learning; y en qué área fueron aplicados: redes, planeamiento de rutas, programación de trabajos y administración de carga. Además, se identificaron ciertas tendencias como que en la mayoría de estas hacían uso del algoritmo DQN y la aplicación más frecuente era en los sectores de redes. En conclusión, se espera que se sigan investigando formas de aplicar estos algoritmos para resolver problemáticas de la sociedad.

Agradecimientos

A mi grupo de apoyo emocional que estuvo en cada paso de la realización de este artículo y a todos aquellos quienes hicieron la realización de esta revisión menos pesada.

Referencias

- [1] L. P. Kaelbling, M. L. Littman, y A. W. Moore, «Reinforcement Learning: A Survey», *Journal of Artificial Intelligence Research*, vol. 4, pp. 237-285, may 1996, doi: 10.1613/jair.301.
- [2] K. Arulkumaran, M. P. Deisenroth, M. Brundage, y A. A. Bharath, «Deep Reinforcement Learning: A Brief Survey», *IEEE Signal Processing Magazine*, vol. 34, n.o 6, pp. 26-38, nov. 2017, doi: 10.1109/MSP.2017.2743240.
- [3] A. Ecoffet, J. Huizinga, J. Lehman, K. O. Stanley, y J. Clune, «Go-Explore: a New Approach for Hard-Exploration Problems», arXiv:1901.10995 [cs, stat], may 2019, Accedido: jul. 18, 2020. [En línea]. Disponible en: <http://arxiv.org/abs/1901.10995>.
- [4] V. Mnih et al., «Human-level control through deep reinforcement learning», *Nature*, vol. 518, n.o 7540, Art. n.o 7540, feb. 2015, doi: 10.1038/nature14236.
- [5] B. Kitchenham, O. Pearl Brereton, D. Budgen, M. Turner, J. Bailey, y S. Linkman, «Systematic literature reviews in software engineering – A systematic literature review», *Information and Software Technology*, vol. 51, n.o 1, pp. 7-15, ene. 2009, doi: 10.1016/j.infsof.2008.09.009.
- [6] I. Khan, X. Tao, G. M. S. Rahman, W. U. Rehman, y T. Salam, «Advanced Energy-Efficient Computation Offloading Using Deep Reinforcement Learning in MTC Edge Computing», *IEEE Access*, vol. 8, pp. 82867-82875, 2020.
- [7] C. Chen, J. Jiang, N. Lv, y S. Li, «An Intelligent Path Planning Scheme of Autonomous Vehicles Platoon Using Deep Reinforcement Learning on Network Edge», *IEEE Access*, vol. 8, pp. 99059-99069, 2020.
- [8] G. Sun, Z. T. Gebrekidan, G. O. Boateng, D. Ayepah-Mensah, y W. Jiang, «Dynamic Reservation and Deep Reinforcement Learning Based Autonomous Resource Slicing for Virtualized Radio Access Networks», *IEEE Access*, vol. 7, pp. 45758-45772, 2019.
- [9] Y. Abiko, T. Saito, D. Ikeda, K. Ohta, T. Mizuno, y H. Mineno, «Flexible Resource Block Allocation to Multiple Slices for Radio Access Network Slicing Using Deep Reinforcement Learning», *IEEE Access*, vol. 8, pp. 68183-68198, 2020.
- [10] Y. Zhang, P. Cai, C. Pan, y S. Zhang, «Multi-Agent Deep Reinforcement Learning-Based Cooperative Spectrum Sensing With Upper Confidence Bound Exploration», *IEEE Access*, vol. 7, pp. 118898-118906, 2019.
- [11] B. Li y Y. Wu, «Path Planning for UAV Ground Target Tracking via Deep Reinforcement Learning», *IEEE Access*, vol. 8, pp. 29064-29074, 2020.

[12]Y. Wei, L. Pan, S. Liu, L. Wu, y X. Meng, «DRL-Scheduling: An Intelligent QoS-Aware Job Scheduling Framework for Applications in Clouds», IEEE Access, vol. 6, pp. 55112-55125, 2018.

[13]Jupiter Bakakeu et al., «An Artificial Intelligence Approach for Online Optimization of Flexible Manufacturing Systems», Applied Mechanics and Materials, vol. 882, pp. 96-108, jul. 2018, doi: 10.4028/www.scientific.net/AMM.882.96.

