

DECIMOQUINTA CONVOCATORIA PARA EL FOMENTO DE LA INVESTIGACIÓN Y LA INNOVACIÓN 2020

Título del proyecto	
Aprendizaje por refuerzo para manipulación de objetos con actuador robótico.	
Campo de acción	Transdisciplinariedad - Aporte al PIM
Desarrollo tecnológico con apuesta social	Este proyecto aporta principalmente a la línea 3 del Plan Integral Multicampus, ya que pretende realizar investigación y desarrollo en un área del conocimiento de vanguardia en las ciencias de la computación, como lo es el aprendizaje por refuerzo, y aplicarlo a un sector altamente productivo, como lo es la robótica. De esta manera, se aporta al desarrollo de sistemas robóticos capaces de mejorar la calidad de vida de las personas, y así mismo se genera conocimiento y producción de alta calidad que permite visibilizar a USTA Colombia en eventos científicos de impacto global.

Articulación con funciones sustantivas y el sector social y productivo

El desarrollo del proyecto en cuestión se articula con las tres funciones sustantivas de la universidad (docencia, investigación y proyección social) de la siguiente manera:

Docencia: El programa de Ingeniería Electrónica de la Universidad Santo Tomás ha planteado, entre otras, dos líneas de profundización: Robótica e Inteligencia Computacional, las cuales al mismo tiempo son líneas de investigación del Grupo de Estudio y Desarrollo en Robótica GED. A partir de esto, en la actualidad se ofrecen dos espacios académicos electivos denominados “Robótica Industrial” e “Introducción a la Inteligencia Artificial”, cuyo objetivo es dotar a los estudiantes de pregrado con las habilidades necesarias para desempeñarse en dichas áreas del conocimiento. El desarrollo del proyecto supone tareas y espera resultados que son acordes a la temática tratada en dichos espacios académicos, lo cual permitiría realizar la articulación de éstos con el proyecto, como proyectos finales conjuntos por parte del joven investigador asociado, o la socialización de los resultados parciales y totales obtenidos en dichos espacios.

Investigación: El proyecto pretende involucrar una temática de vanguardia en el campo de las ciencias de la computación, como es el aprendizaje por refuerzo, en un área de alta importancia en la actualidad, como es la robótica social (aunque los resultados podrían ser extrapolados a otras áreas de la robótica), por lo cual supone una gran importancia y acogida por parte de la comunidad científica. El hecho de llevar planteamientos teóricos a problemas altamente prácticos y tangibles representa una gran oportunidad para dar a conocer sus resultados en eventos científicos de alto impacto, así como en revistas del área,



clasificadas y reconocidas a nivel mundial. De esta manera, se aporta positivamente a los índices y mediciones de investigación de la facultad, la división e incluso la universidad.

Proyección social: Desde varios puntos de vista, el desarrollo en robótica social implica la mejora de la calidad de vida para las personas, algunos ejemplos podrían ser el cuidado de personas de edad o con dificultades motoras, la asistencia a personal de oficina, o incluso el apoyo en las tareas domésticas. El desarrollo de este tipo de tecnología impactará a un gran número de personas de forma positiva, por lo cual es pertinente que la universidad esté involucrada en dicha evolución tecnológica.

Sector productivo: La robótica en general está cada día más presente en la industria colombiana, por lo cual involucrar a los futuros profesionales que la universidad forma, en este tipo de tecnología, es crucial para estar a la vanguardia.

Grupo de investigación	Línea de investigación en la que se inscribe el proyecto
Grupo de Estudio y Desarrollo en Robótica GED	Robótica Inteligencia Computacional

Nombre del Investigador principal	Enlace CvLAC	Enlace ORCID	Enlace Google Académico
Edgar Camilo Camacho Poveda	http://scienti.colciencias.gov.co:8081/cvllac/visualizador/generarCurriculoCv.do?cod_rh=0001630084	https://orcid.org/0000-0002-6084-2512	https://scholar.google.com/citations?user=tJG988kAAAAJ&hl=es
División	Facultad	Programa	Grupo de investigación
Ingeniería	Ingeniería Electrónica	Ingeniería Electrónica	Grupo de Estudio y Desarrollo GED
Nombre del Co-investigador	Enlace CvLAC	Enlace ORCID	Enlace Google Académico
Fabian Eduardo Pérez Gordillo	http://scienti.colciencias.gov.co:8081/cvllac/visualizador/generarCurriculoCv.do?cod_rh=0001516111	https://orcid.org/0000-0002-2746-8733	https://scholar.google.com/citations?user=vncSAb0AAAAJ&hl=es



División	Facultad	Programa	Grupo de investigación
Ingeniería	Ingeniería Electrónica	Ingeniería Electrónica	Grupo de Estudio y Desarrollo GED

Resumen de la propuesta	Palabras clave
Este proyecto implementará un algoritmo de aprendizaje por refuerzo para la manipulación de objetos por parte de un brazo robótico, enfocado a tareas que deba realizar un robot de servicio doméstico. Partirá de la implementación de la plataforma robótica, la cual tendrá los grados de libertad y el actuador adecuados para la manipulación de objetos encontrados en situaciones cotidianas. A continuación, se planteará el algoritmo de aprendizaje por refuerzo, partiendo del espacio de estados que mejor describa al ambiente, y el espacio de acciones que permita realizar las tareas de forma adecuada. Así mismo se estudiarán métodos que permitan aproximar el estado a partir de la observación directa el ambiente, como redes neuronales profundas. El entrenamiento se realizará en su mayoría en un entorno de simulación, con el fin de no poner en riesgo la integridad física del robot en las primeras etapas de aprendizaje. Al final del proceso, si se considera necesario, se refinará el entrenamiento obtenido con el robot real. A continuación, se evaluará cuantitativamente el resultado obtenido por el agente en las diferentes tareas aprendidas. Finalmente, se iniciará con la implementación de este brazo robótico en un prototipo de robot social doméstico construido en su totalidad por la Universidad Santo Tomás, con el fin de participar en certámenes de robótica social y doméstica.	Aprendizaje por refuerzo Aprendizaje de máquina Robótica social Manipulación de objetos

Problema de investigación

Día a día aumenta el número de personas que necesita algún tipo de apoyo en sus labores cotidianas, ya sea por su edad o por problemas de salud específicos, por ejemplo, se estima que en la unión europea hay alrededor de 9 millones de personas que necesitan ayuda para levantarse de la cama [1]. En el contexto nacional, el Ministerio de Salud y Protección Social estima que para el año 2020, habrá 50 adultos mayores por cada 100 adolescentes en Colombia [2]. Hace unos años se hablaba constantemente de la disminución de la tasa de natalidad en países industrializados [3], lo cual provocó que en la actualidad se evidencien altos números de personas de edad avanzada en contraste a la cantidad de personas jóvenes. Dicho tema ahora está presente en Colombia, donde se evidencian reducciones en las tasas de natalidad [4], a partir de lo cual se podría inferir que en unos años se tendrán proporciones desfavorables de jóvenes respecto a personas de edad avanzada.

La presencia de este problema a nivel mundial ha ocasionado que la comunidad científica enfoque gran parte de su esfuerzo en desarrollar sistemas que permitan apoyar a las personas de edad avanzada en tareas domésticas, llegando a la conclusión que el área que más puede impactar de forma positiva es la robótica, y más específicamente, la robótica social. Por esta razón, las hojas de ruta de la robótica más actuales para Europa [1] y Estados Unidos [5] plantean el desarrollo de la robótica social como una herramienta esencial para afrontar el problema planteado, con el fin de ofrecer autonomía a todas las personas que la puedan necesitar, en el caso en que las proporciones poblacionales no permitan que una persona con todas sus capacidades las pueda apoyar en todo momento.

Del problema planteado anteriormente, se desencadena un problema más específico desde el punto de vista técnico, y es la manipulación de objetos por parte de dichos robots sociales. Entre las tareas más comunes de asistencia robótica está el manipular objetos con el fin de alcanzarlos al humano, como bebidas, comidas, etc., o transportarlos a los lugares adecuados, en los casos de sacar la basura, o simplemente utilizarlos como herramientas para otras tareas, como regar plantas. Siempre que la tarea esté especificada de fábrica, matemáticamente será posible realizar los movimientos necesarios para ejecutarla, sin embargo, muchas tareas se deben aprender según las diferentes situaciones a las que el robot se enfrenta. Este proyecto se enfoca en el problema de aprender a manipular nuevos objetos (desconocidos para el robot), en condiciones espaciales igualmente nuevas (por ejemplo, obstáculos dinámicos). Se plantea utilizar aprendizaje por refuerzo, con el fin de que el robot pueda aprender de su propia experiencia, y mejorar su técnica a medida que practica por sí mismo.

La problemática anteriormente planteada da cabida a la siguiente pregunta de investigación:

¿Cómo dotar a un robot con la capacidad de aprender a manipular objetos a partir de su propia experiencia, sin la necesidad de programar explícitamente estas acciones de forma previa?

Justificación

En el problema anteriormente planteado se encontró la necesidad de incrementar el grado de independencia que las personas de edad avanzada o con problemas de movilidad pueden tener, y se presentó como alternativa la robótica social. El estudio y desarrollo en esta área acerca cada vez más el escenario en que los robots domésticos puedan ayudar con las tareas básicas que una persona en las condiciones planteadas necesita, mejorando altamente su calidad de vida. El impacto social que los robots de asistencia social tendrán será de gran magnitud, ya que permitirán descongestionar centros médicos y posibilitarán que las personas de edad avanzada puedan convivir cerca de sus familiares [6].

La implementación de aprendizaje por refuerzo (RL por sus siglas en inglés) para la manipulación de objetos permite que, en la etapa de desarrollo de un sistema robótico social, no se tengan que plantear todas las tareas que el robot deba realizar (las cuales posiblemente no se podrán contemplar en su totalidad), sino que éste puede aprender a realizarlas en el ambiente y bajo las condiciones de trabajo finales. En el caso específico de este proyecto, el RL para manipulación de objetos abrirá la posibilidad que el robot aprenda a sujetar, transportar y liberar diferentes y nuevos objetos para los que no fue programado inicialmente, aumentando la productividad y mejorando la calidad de la ayuda que puede prestar.

Además, el estudio de una nueva tecnología, como lo es el RL, y su aplicación en un tema tan transversal como la robótica, permitirá realizar desarrollos que incrementen la visibilidad de la Universidad Santo Tomás ante la comunidad científica como un centro de investigación que está a la vanguardia de la inteligencia artificial y la robótica, a través de la participación en eventos internacionales y la publicación en revistas de alto impacto a nivel global. Los desarrollos generados en este proyecto permitirán además iniciar con el nuevo proyecto del Grupo de Estudio y Desarrollo en Robótica, que pretende diseñar un sistema de asistencia doméstica completamente de autoría de la Universidad Santo Tomás, con el fin de participar en la Competencia Latinoamericana de Robótica en 2020 en su liga de robótica social, y posteriormente (2021 o después) en la competencia mundial RoboCup, en la liga @home open. Estas competencias permitirán dar un gran nivel de visibilidad de la universidad a nivel mundial.

Objetivo general

Implementar un algoritmo de aprendizaje por refuerzo, a través del planteamiento de la arquitectura, el refinamiento y el entrenamiento adecuado, con el fin de brindar a un robot la capacidad de aprender a realizar nuevas tareas de manipulación partiendo de su propia experimentación.

Objetivos específicos

- Definir un conjunto de estados y acciones apropiadas para describir el ambiente en que funcionará el brazo robótico.
- Plantear escenarios de prueba, cuyo objetivo sea la manipulación de diferentes tipos de objetos.
- Entrenar el algoritmo de aprendizaje por refuerzo en ambientes simulados y/o reales.
- Evaluar el desempeño del algoritmo en escenarios reales, a partir de objetivo de manipulación planteado.

Estado del arte y marco conceptual

Marco conceptual: Aprendizaje por refuerzo y *Q-Learning*

El aprendizaje por refuerzo (RL por sus siglas en inglés) está planteado bajo la idea de que un agente cree una política que le permita tomar las acciones adecuadas según el estado que observe del ambiente, con el fin de maximizar una recompensa predefinida a largo plazo. Dicha política es creada a partir de la experimentación del agente, por lo cual se podría decir que éste aprende a partir de prueba y error. El ambiente se puede describir como un Proceso de Decisión de Markov (MDP por sus siglas en inglés), en el cual el agente recibe una observación (estado) del ambiente s_t en cada tiempo discreto t , y toma una acción a_t correspondiente a su política. La acción afecta al ambiente, el cual entrega al agente el nuevo estado s_{t+1} y una recompensa numérica r_t que representa el beneficio de haber tomado dicha acción en dicho estado (véase la Figura 1) [7].

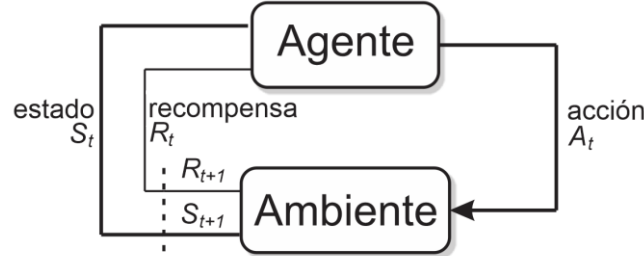


Figura 1. Diagrama de bloques del Aprendizaje por Refuerzo

El agente tiene como objetivo aprender a maximizar las recompensas que obtiene a largo plazo, planteando una función $Q(s_t, a_t)$, que mapea la utilidad esperada de tomar la acción a_t en el estado s_t . A partir del valor Q obtenido, el agente genera una política $\pi_t(s_t, a_t)$, entrega la probabilidad de aplicar la acción a_t en el estado s_t , que asegure la máxima recompensa a largo plazo. Los métodos de aprendizaje por refuerzo varían principalmente en la forma en que plantean o actualizan la función de valor Q . *Q-Learning* utiliza la siguiente formula o regla de actualización:

$$Q(s^t, a^t) = (1 - \alpha)Q(s^t, a^t) + \alpha \left[r + \gamma \max_{a \in A} Q(s^{t+1}, a) \right]$$

Dónde r corresponde a la recompensa obtenida en el tiempo t ($r \in \mathbb{R}$), α es la tasa de aprendizaje en el tiempo t ($0 \leq \alpha \leq 1$), y γ es la tasa de descuento ($0 \leq \gamma \leq 1$). La actualización se realiza a partir de la acción más ambiciosa (*greedy actions*) sin tener en cuenta la política actual. Así mismo, la acción tomada por el agente corresponderá, en principio, a la *greedy action*, lo que se denomina *explotar* el aprendizaje adquirido hasta el momento:

$$A_t = \underset{a}{\operatorname{argmax}} Q_t(s^t, a^t)$$

Con el fin que el agente pueda aprender constantemente, no siempre es conveniente *explotar* la acción con mayor valor, en su lugar, es necesario que se *exploren* acciones diferentes con el fin de conocer su impacto en el ambiente, en este caso se utiliza selección de acción ε – *greedy*, que corresponde a tomar una acción aleatoria con probabilidad ε , y la acción *greedy* con probabilidad $(1 - \varepsilon)$ [8].

Estado del arte

Desde el punto de vista del RL, la manipulación usando actuadores robóticos presenta un problema de alta dimensionalidad, tanto en el espacio del estado como el de la acción. Además, la obtención de datos es costosa desde el punto de vista temporal, debido a que un robot real se toma algunos segundos o incluso minutos para realizar una tarea específica. En muchos casos, es necesario modelar el sistema como un MDP parcialmente observable, o incluso no observable [9]. Estos y otros inconvenientes han ocasionado que el problema en cuestión se esté trabajando desde diferentes variaciones de algoritmos de aprendizaje por refuerzo [10].

El problema general de aprendizaje por refuerzo aplicado a problemas con altas dimensiones se puede afrontar desde diferentes técnicas. Partiendo de redes profundas que reemplazan a la tabla Q, o *Deep Q-Nerwoks*, presentándose en [11] una de sus primeras aplicaciones prácticas con técnica ε – *greedy* para la exploración. Dicha técnica de explotación puede volverse adaptativa como en [12], donde ε se ajusta según la diferencia temporal. Otras técnicas que no se basan en tabla de valor (tabla Q), sino que modifican directamente la política, como *Trust Region Policy Optimization (TRPO)* [13] y *Proximal Policy Optimization (PPO)* [14], exploran a partir de la última versión de la política. [15] presenta una implementación de *Deep Q-Learning* para la generación de políticas en el dominio continuo. Otros métodos de exploración incluyen regularización de entropía [16], *adversarial self-play* [17] o *Parameter Space Noise* [18].

Para las aplicaciones específicas de manipulación de objetos y control de manipuladores robóticos, se plantean diferentes técnicas de aprendizaje por refuerzo. [19] plantea diferentes técnicas para la manipulación de brazos robóticos de 2 y 3 grados de libertad. [24] por su lado aprovecha la simetría de un robot de dos brazos robóticos. El *curriculum learning*, usado en [20], parte de la idea de enseñar al agente la tardea a partir de subtarefas de menor dificultad, y a medida que el agente “apruebe” cada lección, se le va asignando una más complicada hasta llegar a la deseada. En el caso de aprendizaje por refuerzo inverso, se pretende que el agente infiera el objetivo o la recompensa a optimizar, a partir de la observación de las trayectorias que otro agente “experto” toma, [21] lo implementa utilizando *Q-Learning*.

Los anteriores trabajos basan su actuar y explorar en el valor dado a cada acción para los posibles estados (valor Q), sin embargo, otras metodologías se basan directamente en la política. El trabajo presentado en [22] corresponde a este caso, donde se realiza un aprendizaje en lazo cerrado. Por otro lado, [23] utiliza aprendizaje profundo por refuerzo (*Deep Reinforcement Learning DRL*), utilizando una redes P profundas (*Deep P-Networks*) para la manipulación de ropa con brazos robóticos. Por otro lado, en [23] se aplica *PPO* y *DRL* para labores de recoger y colocar (*pick and place*) en aplicaciones de logística.

Metodología

Para el desarrollo de este proyecto se propone un conjunto de etapas a llevar a cabo de forma no secuencial, debido a que los resultados obtenidos en algunas de ellas pueden provocar la continuidad, o la repetición de otras. La Figura 2.

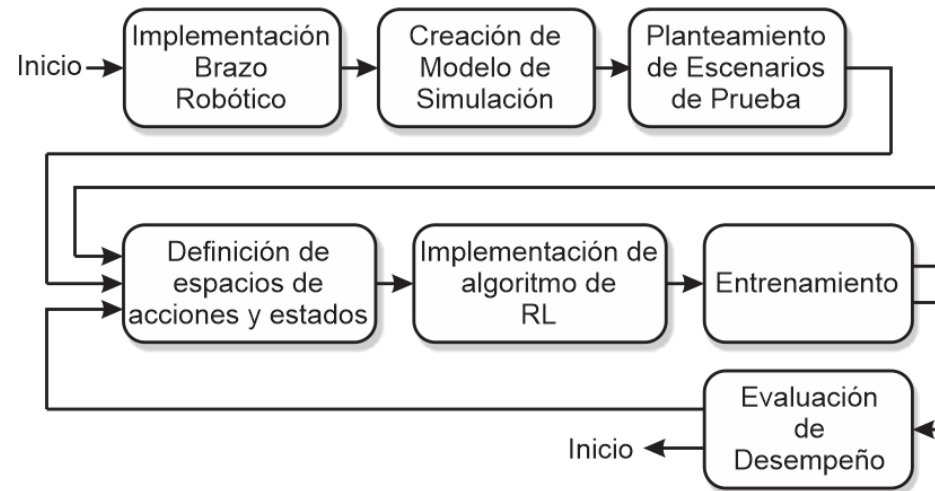


Figura 2. Flujo del desarrollo de las diferentes etapas del proyecto.

Implementación del brazo robótico: Se adquirirán los materiales necesarios para la construcción de un brazo robótico con el actuador adecuado, para manipular diferentes tipos de objetos que se encuentren en situaciones cotidianas. Además, se implementarán las tarjetas electrónicas y de control, y las interfaces para su comunicación con el computador. A continuación, se desarrollarán las librerías, controladores y demás elementos de software necesarios para controlar el brazo desde el sistema operativo para robótica ROS.

Creación de modelo de simulación: Después de tener construido el brazo, se crearán sus modelos 3D, el archivo de descripción unificada del robot (URDF), y los controladores necesarios para la simulación en el software especializado Gazebo. A continuación, se crearán los nodos necesarios para su comunicación con ROS.

Planteamiento de escenarios de prueba: Se evaluarán las situaciones a las que un robot de servicio doméstico se pueda afrontar, partiendo principalmente de las pruebas estandarizadas que se evalúan en la competencia RoboCup@Home [26], entre las cuales se encuentra la manipulación de desechos, limpiar superficies, servir bebidas, entre otras. A partir de dichas pruebas, se creará una población de los posibles objetos a manipular, junto con el contexto correspondiente. A continuación, estudiando las capacidades del laboratorio de robótica y de las instalaciones de la universidad en general, se creará la

muestra sobre la cual se aplicará el estudio, correspondiente a un conjunto específico de pruebas estandarizadas, donde se establecerán los objetos a manipular y los escenarios que

Definición de espacios de estados y acciones: Como el algoritmo de aprendizaje por refuerzo entrenará a un agente para la realización de las tareas de manipulación, es necesario definir el espacio de estados que describirá todas las posibles situaciones del ambiente, así como las posibles acciones éste podrá tomar. Sin embargo, según la complejidad del ambiente observado, algunas veces es complejo plantear la observación del ambiente de forma analítica, ya sea discreta o continua, por lo cual las últimas generaciones de algoritmos de RL plantean la utilización de estimadores de estado, que toman observaciones directas del ambiente (por ejemplo información visual), y a partir de métodos de aprendizaje de máquina crean modelos para estimar el estado [27] [11].

Implementación del algoritmo de aprendizaje por refuerzo: Se evaluarán diferentes técnicas y arquitecturas para el aprendizaje por refuerzo, escogiendo así la que mejor resultados presente en las pruebas preliminares. Se estudiarán diferentes técnicas como *Q-Learning*, *Proximal Policy Optimization (PPO)*, *Actor Critic*, entre otros. Así mismo se especificará el conjunto de parámetros en cada caso, y sus correspondientes barridos para el entrenamiento. En adición, se aplicarán diferentes métodos de aproximación de estado como redes neuronales, redes convolucionales, redes recurrentes, entre otras. Se migrará dicha implementación a ROS, con el fin de comunicarse con el modelo en simulación y/o con el robot real.

Entrenamiento: Se ejecutarán los entrenamientos, inicialmente en simulación, con el fin de obtener políticas aproximadas del comportamiento deseado, sin poner en riesgo la integridad del robot real. A continuación, si se considera pertinente, se entrenará con el robot real con el fin de refinar los resultados obtenidos.

Evaluación de desempeño: El instrumento para la evaluación numérica del desempeño del modelo obtenido se basará en las pruebas estandarizadas planteadas en conjunto con los escenarios, que permitan la comparación cuantitativa de los diferentes métodos a partir de la ejecución de las tareas aprendidas por parte del agente que controlaría al robot.

De forma paralela al desarrollo del algoritmo de aprendizaje, el brazo robótico será una parte de un proyecto macro del Grupo de Estudio y Desarrollo en Robótica de la Universidad Santo Tomás, correspondiente al desarrollo e implementación de un robot de servicio doméstico. Por esta razón, se trabajará en la construcción de un primer prototipo de este sistema, el cual será evaluado en una competencia internacional, correspondiente a la Competencia Latinoamericana de Robótica en 2020 en su liga de robótica social, que se realizará en Brasil. El objetivo a largo plazo de este proyecto es participar en competencia mundial RoboCup, en la liga @home open.

Resultados esperados

A partir del desarrollo de este proyecto, se esperan los siguientes productos:

- Nuevo conocimiento: Un (1) artículo Q3 o Q4.
- Diseño Tecnológico e Innovación: Un (1) registro de software.
- Producto de apropiación social del conocimiento: Una (1) participación en evento científico.
- Producto de formación de recurso humano: Dirección de dos (2) trabajos de grado de pregrado.

Cronograma

ACTIVIDAD	RESPONSABLE	FECHA		FEBRERO				MARZO				ABRIL				MAYO				JUNIO				JULIO				AGOSTO				SEPTIEMBRE				OCTUBRE				NOVIEMBRE			
		Inicio	Fin.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	28	29	30	31	32	33	34	35	36	37	38	39	40	41
Implementación de brazo robótico	Camilo Camacho Fabian Pérez	02/2020	03/2020																																								
Creación de modelo de simulación del robot construido	Camilo Camacho Fabian Pérez	03/2020	04/2020																																								
Definición espacios de estados y acciones	Camilo Camacho	05/2020	05/2020																																								
Implementación de algoritmo de aprendizaje por refuerzo	Camilo Camacho	06/2020	06/2020																																								
Entrenamiento de algoritmo de aprendizaje por refuerzo	Camilo Camacho	07/2020	08/2020																																								
Evaluación de desempeño de algoritmos entrenados	Camilo Camacho	09/2020	09/2020																																								
Inicio de la implementación de robot de	Camilo Camacho Fabian Pérez	04/2020	11/2020																																								

[illegible]

Presupuesto					
Horas nómina					
Concepto	Nombre	Escalafón	Horas mes	Sede / Seccional o Externo	Total (\$)
Horas Nomina (Investigador Principal)	Edgar Camilo Camacho Poveda	2	35	USTA Sede Principal	\$9'386.562
Horas Nomina (Co-Investigadores)	Fabian Eduardo Pérez Gordillo	2	29	USTA Sede Principal	\$7'777.437



FINANCIACIÓN	RECURSO	DESCRIPCIÓN	Valor partida	Valor contrapartida (Externa)	Total (\$)
RUBROS	Servicios Técnicos	N/A	\$ 0	\$ 0	\$ 0
	Salidas de campo	N/A	\$ 0	\$ 0	\$ 0
	Equipos	N/A	\$ 0	\$ 0	\$ 0
	Materiales, insumos y software	• Materiales para desarrollo de brazo robótico. • Materiales para el inicio del desarrollo de robot de servicio doméstico. • GPU para entrenamiento de algoritmos de inteligencia artificial y aprendizaje de máquina. • Materiales mecánicos y electrónicos en general.	\$ 0	\$ 0	\$ 10'000.000
BOLSAS	Papelería	N/A	\$ 0	\$ 0	\$ 0
	Fotocopias	N/A	\$ 0	\$ 0	\$ 0
	Material bibliográfico	N/A	\$ 0	\$ 0	
	Auxilio de transporte	N/A	\$ 0	\$ 0	\$ 0
	Movilidad	• Asistencia a conferencia internacional. • (Opcional) Asistencia a la competencia latinoamericana de robótica en 2020 en Brasil, con dos estudiantes de pregrado.	\$ 0	\$ 0	\$ 6'000.000
	Publicaciones (Artículos, proceso editorial y traducción)	Costos para publicación de un artículo en revista Q3 o Q4.	\$ 0	\$ 0	\$ 0
TOTAL DEL PROYECTO:					\$ 16'000.000

Referencias bibliográficas

- [1] SPARC. The Partnership for Robotics in Europe, Robotics 2020 Multi-Annual Roadmap For Robotics in Europe, Europe: – Horizon 2020, 2015.
- [2] Ministerio de Salud y Protección Social, «Envejecimiento demográfico. colombia 1951-2020 dinámica demográfica y estructuras poblacionales,» Bogotá DC, 2013.
- [3] G. Nargund, «Declining birth rate in Developed Countries: A radical policy re-think is required,» *Facts Views Vis Obgyn*, vol. 1, p. 191–193, 2009.
- [4] Revista Dinero, «Cada vez nacen menos bebés en Colombia, ¿hacia dónde vamos?,» 22 11 2018. [En línea]. Available: <https://www.dinero.com/edicion-impresa/pais/articulo/tasa-de-natalidad-en-colombia-disminuye-cada-vez-mas/264428>. [Último acceso: 28 07 2019].
- [5] University of California San Diego, Carnegie Mellon University, others, A Roadmap for US Robotics. From Internet to Robotics. 2016 Edition, USA, 2016.
- [6] E. B. a. B. M. Tony Kuo, «Designing a robotic assistant for healthcare applications,» *Robotics in Healthcare Applications*, 2018.
- [7] R. S. Sutton y A. G. Barto, Reinforcement Learning:, London, England: The MIT Press, 2018.
- [8] C. Watkins, «Learning From Delayed Rewards,» King's College, Cambridge, UK, 1989.
- [9] H. M. L. Hai Nguyen, «Review of Deep Reinforcement Learning for Robot Manipulation,» *2019 Third IEEE International Conference on Robotic Computing (IRC)*, 2019.
- [10] S. Kelly y M. I. Heywood, «Emergent Solutions to High-Dimensional Multitask Reinforcement Learning,» *Evolutionary Computation*, vol. 26, nº 3, pp. 347-380, 2018.
- [11] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra y M. Riedmiller, «Playing Atari with Deep Reinforcement Learning,» de *NIPS Deep Learning Workshop 2013*, Toronto, Canadá, 2013.
- [12] M. Tokic12, «Adaptive ϵ -Greedy Exploration in Reinforcement Learning Based on Value Differences,» Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.
- [13] J. Schulman, S. Levine, P. Moritz, M. I. Jordan y P. Abbeel, «Trust Region Policy Optimization,» de *Intern. Conf. on Machine Learning*, 2015.
- [14] J. Schulman, F. Wolski, P. Dhariwal, A. Radford y O. Klimov, «Proximal policy optimization algorithms,» *Axiv*, vol. abs/1707.06347, 2017.

- [15] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver y D. Wierstra, «Continuous control with deep reinforcement learning,» *ArXiv:1509.02971*, n° arXiv:1509.02971, 2015.
- [16] T. Haarnoja, A. Zhou, P. Abbeel y S. Levine, «Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor,» *ArXiv*, n° arXiv:1801.01290.
- [17] S. Sukhbaatar, Z. Lin, I. Kostrikov, G. Synnaeve, A. Szlam y R. Fergus, «Intrinsic Motivation and Automatic Curricula via Asymmetric Self-Play,» de *Sixth International Conference on Learning Representations*, Montreal, Canadá, 2018.
- [18] M. Plappert, R. Houthooft, P. Dhariwal, S. Sidor, R. Y. Chen, X. Chen, T. Asfour, P. Abbeel y M. Andrychowicz, «Parameter Space Noise for Exploration,» de *Seventh International Conference on Learning Representations*, Montreal, Canadá, 2018.
- [19] J. J. Sweafford y F. Fahimi, «MODEL-FREE ONLINE REINFORCEMENT LEARNING OF A ROBOTIC MANIPULATOR,» *MECHATRONIC SYSTEMS AND CONTROL*, vol. 47, n° 3, pp. 136-143, 2019.
- [20] M. Breyer, F. Furrer, T. Novkovic, R. Siegwart y J. Nieto, «Comparing Task Simplifications to Learn Closed-Loop Object Picking Using Deep Reinforcement Learning,» *IEEE ROBOTICS AND AUTOMATION LETTERS*, vol. 4, n° 2, pp. 1549-1556, 2019.
- [21] S. Krishnan, A. Garg, R. Liaw, B. Thananjeyan, L. Miller, F. T. Pokorny y K. Goldberg, «SWIRL: A sequential windowed inverse reinforcement learning algorithm for robot tasks with delayed rewards,» *INTERNATIONAL JOURNAL OF ROBOTICS RESEARCH*, vol. 38, n° 2-3, pp. 126-145, 2019.
- [22] T. G. Thrun, E. Falotico, F. Renda y C. Laschi, «Model-Based Reinforcement Learning for Closed-Loop Dynamic Control of Soft Robotic Manipulators,» *IEEE TRANSACTIONS ON ROBOTICS*, vol. 35, n° 1, pp. 124-134, 2019.
- [23] Y. Tsurumine, Y. Cui, E. Uchibe y T. Matsubara, «Deep reinforcement learning with smooth policy update: Application to robotic cloth manipulation,» *ROBOTICS AND AUTONOMOUS SYSTEMS*, vol. 112, pp. 72-83, 2018.
- [24] F. Amadio, A. Colome y C. Torras, «Exploiting Symmetries in Reinforcement Learning of Bimanual Robotic Tasks,» *IEEE ROBOTICS AND AUTOMATION LETTERS*, vol. 4, n° 2, pp. 1938-1845, 2019.
- [25] A. Iriondo, E. Lazkano, L. Susperregi, J. Urain y A. Fernandez, «Pick and Place Operations in Logistics Using a Mobile Manipulator Controlled with Deep Reinforcement Learning,» *APPLIED SCIENCES-BASEL*, vol. 9, n° 2, p. 348, 2019.
- [26] M. Matamoros, C. Rascon, S. Wachsmuth, A. Moriarty, J. Kummert, J. Hart, S. Pfeiffer, M. van der Brugh y M. St-Pierre, RoboCup@Home, Rules & Regulations, 2019.

- [27] H. Li, R. Cai, N. Liu, X. Lin y Y. Wang, «Deep reinforcement learning: Algorithm, applications, and ultra-low-power implementation,» *NANO COMMUNICATION NETWORKS*, vol. 16, pp. 81-90, 2018.

