PLAN DE INVESTIGACIÓN

|  |  |
| --- | --- |
| **DATOS DEL DOCTORANDO** | |
| **Datos personales** | |
| Apellidos | Nombre |
| Lucas Zaragoza | Rubén |
| DNI, NIE, Pasaporte | Email |
| 06282454G | [ruben.lucas.zaragoza@gmail.com](mailto:ruben.lucas.zaragoza@gmail.com) |
| **Datos Académicos** | |
| Programa de Doctorado en | Linea de investigación |
| Tecnologías de la Información y las  Comunicaciones | Inteligencia artificial, aprendizaje por refuerzo, Aplicación práctica de  Deep reinforcement learning en robots |
| Título de la tesis | |
|  | |

|  |  |
| --- | --- |
| **DATOS DE DIRECCIÓN DE TESIS** | |
| **Director de Tesis** |  |
| Apellidos | Cañas Plaza |
| Nombre | José María |
| DNI |  |
| Email | [josemaria.plaza@urjc.es](mailto:josemaria.plaza@urjc.es) |
| Departamento/Universidad/Centro | Teoría de la Señal y Comunicaciones y Sistemas  Telemáticos/Universidad Rey Juan Carlos |
| **Codirector de Tesis** |  |
| Apellidos |  |
| Nombre |  |
| DNI |  |
| Email |  |
| Departamento/Universidad/ Centro |  |
| **Tutor (si difiere del Director de Tesis)** |  |
| Apellidos |  |
| Nombre |  |
| DNI |  |
| Email |  |
| Departamento/Universidad/ Centro |  |

Los datos personales recogidos serán incorporados y tratados en el fichero *Tesis Doctorales*, cuya finalidad es la gestión de las tesis doctorales, inscrito en el Registro de Ficheros de Datos Personales de la Agencia Española de Protección de Datos (https://[www.agpd.es/portalwebAGPD/index-ides-idphp.php).](http://www.agpd.es/portalwebAGPD/index-ides-idphp.php)) El órgano responsable del fichero es el Vicerrectorado de Títulos Propios, Formación Continua y Posgrado, y la dirección donde el interesado podrá ejercer los derechos de acceso, rectificación, cancelación y oposición ante el mismo es, Vicerrectorado de Títulos Propios, Formación Continua y Posgrado, C/ Tulipán s/n, 28933-Móstoles, todo lo cual se informa en cumplimiento del artículo 5 de la Ley Orgánica 15/1999, de 13 de diciembre, de Protección de Datos de Carácter Personal.

**DATOS DE LA TESIS**

Durante la elaboración de esta tesis doctoral se pretende explorar diversas técnicas de machine learning que permitirán a un robot completar tareas de diversa índole de manera satisfactoria.

La investigación se centrará en concreto en la implementación de algoritmos de reinforcement learning, Deep reinforcement learning y Deep learning.

En una primera instancia se resolverán una serie de problemas básicos en un simulador que nos permitirá la realización de una evaluación objetiva en un entorno más controlado para, posteriormente, ser capaces de aplicar estas técnicas en un problema más complejo que se pueda llevar a un entorno real con un robot físico. Durante la realización de las pruebas se colaborará con el desarrollo de RL-Studio, lo cual permitirá aplicar distintos algoritmos de inteligencia artificial a robots en un entorno simulado.

La versión utilizada se compondrá de:

* Gazebo 11 y Carla 0.9.12 para simular el entorno, los robots y las físicas que aplicarán al problema
* Ros noetic para comunicarse con los robots
* Librerías que nos permitirán implementar los algoritmos con los que se va a trabajar en esta tesis.
* Integración con Behavior metrics para evaluar y comparar el rendimiento de los agentes entrenados

1. Resumen del proyecto de tesis

2. Justificación del tema de investigación. Antecedentes y estado actual del tema.

La aplicación de inteligencia artificial para resolver problemas concretos está creciendo tanto en la empresa privada como en la empresa pública en todos los sectores (medicina, finanzas, industria, automoción, etc.).

Esto se debe a la utilidad que demuestra tener para mejorar la calidad de productos y servicios y la disminución de los costes asociados a estos, aumentando no solo los beneficios que las empresas obtienen, sino también la calidad de vida de las personas.

Dentro de este ámbito se encuentra el aprendizaje por refuerzo, el cual es capaz de obtener resultados de una manera similar a como lo hace el propio ser humano, aprendiendo del entorno y sin necesidad de humano que provea de información pormenorizada al sistema en cuestión.

Esto se extiende también a los robots, los cuales, a base de ensayo y error, son capaces de aprender lo que se conoce como una política óptima para la consecución de los objetivos que se le plantean. Haciendo posible que situaciones que antes requerían de un esfuerzo mayor por parte de una o varias personas se pueda programar en un robot para facilitar tareas o incluso ejecutarlas de manera más precisa y eficiente.

Una de estas tareas en las que la comunidad lleva tiempo aunando esfuerzos es la conducción autónoma, ámbito en el cual se desarrollará la tesis.

La conducción autónoma ha experimentado avances significativos en los últimos años, acercándonos a un futuro en el que los vehículos autónomos sean comunes en nuestras carreteras. Numerosas empresas y estudios de investigación han logrado un progreso notable en el desarrollo de tecnologías de conducción autónoma, pero aún existen desafíos en algunas áreas.

A continuación se enumeran una serie de desafíos que podrían ser abordados:

Capacidad de coordinar conducción con otros vehículos para reducir atascos

Capacidad de realizar una conducción eficiente teniéndo en cuenta inputs como las condiciones de la via o el consumo de gasolina para determinadas maniobras

Conducción segura en condiciones climatológicas adversas.

* Maximizar el nivel de seguridad en la conducción autónoma: Uno de los problemas principales que aún no se han resuelto por completo es lograr un alto nivel de seguridad y confiabilidad. Si bien los vehículos autónomos han mostrado resultados prometedores en entornos controlados, navegar por escenarios del mundo real complejos sigue siendo un desafío considerable. El comportamiento humano impredecible, las condiciones climáticas adversas y las carreteras mal señalizadas pueden representar obstáculos significativos para los vehículos autónomos hasta el punto de que empresas como Waymo, Tesla y Cruise aún se encuentran en fase de desarrollo para poder garantizar una solución segura.
* Cubrir casos de uso adversos mejorando la percepción. Otro desafío importante es garantizar una percepción y comprensión robustas del entorno circundante. Los vehículos autónomos se basan en una combinación de sensores, como cámaras, lidar, radar y GPS, para percibir e interpretar de manera precisa su entorno. Sin embargo, estos sensores pueden verse afectados por condiciones climáticas adversas, como lluvia intensa o niebla, lo que provoca un rendimiento degradado. Para abordar esto, los investigadores están explorando técnicas de fusión de sensores que integran datos de múltiples sensores para mejorar las capacidades de percepción en diversas condiciones.
* Optimizar la conducción: Otra linea de investigación en ciernes es la posibilidad de intercomunicar los diferentes vehículos para, prediciendo la secuencia de acciones que el resto de agentes puede tomar, reducir atascos y evitar ineficiencias en la conducción. A esto se podrían añadir distintos inputs como el estado de la via o el consumo de combustible para no solo reducir ineficiencias, si no también evitar posibles accidentes y reducir la contaminación.

En cuanto a los avances, los algoritmos de Aprendizaje por Refuerzo Profundo como DQN, PPO o DDPG han demostrado ser prometedores para mejorar las capacidades de conducción autónoma. Esto, unido a otras técnicas complementarias como imitation learning o extractores de características neuronales han posibilitado que distintos tipos de agentes ya sean capaz de resolver multitud de retos de manera individualizada, entre los que podemos mencionar negociación de cruces, resolución de circuitos predefinidos con tráfico y peatones o transferencia de conocimiento de agentes simulados a vehículos en el mundo real.

3. Hipótesis de trabajo y Objetivos

En este proyecto se parte de las siguientes hipótesis:

1. dando al robot las reglas que rigen el entorno, este será capaz de aprender a lograr el objetivo para el que se diseña. Es decir, dada una serie de estados por los que pasará el robot a lo largo del problema y una serie de acciones que debe acometer para lograr un objetivo, este será capaz de aprender cuales son las acciones óptimas que debe acometer en función de los estados en los que se encuentre, los cuales serán captados a través de sensores de distinta naturaleza.
2. Es posible transferir el conocimiento adquirido en un simulador realista a un entorno real.

Los objetivos a alcanzar son los siguientes:

1. Resolver problemas básicos de reinforcement learning clásico en un simulador lo más realista posible.
2. Abordar y resolver problemas complejos aplicando técnicas de aprendizaje por refuerzo más sofisticadas como Deep reinforcement learning con distintos tipos de robots (coche, roomba, dron, etc.) de manera que podamos abstraer lo máximo posible la implementación de los algoritmos de la capa de control de cada robot (incluyendo sensorización de estados y aplicación de acciones).
3. Contribuir al desarrollo de una herramienta que permita adaptar de manera sencilla la aplicación de estos algoritmos a distintos problemas configurando los distintos tipos de sensores, las distintas acciones, el objetivo que se persigue y la naturaleza del problema. De esta manera, un futuro usuario de la herramienta podrá abstraerse de la implementación de estos algoritmos y resolver distintos problemas de manera rápida y efectiva.
4. Abordar un problema práctico aún no resuelto por la comunidad investigadora o con cierto margen de mejora.
5. Contribuir al estado del arte no solo en el ámbito teórico, si no en el ámbito práctico, resolviendo este problema en un entorno real de manera satisfactoria.

4. Metodología

La metodología que se emplea a lo largo de esta investigación se basa en la realización de experimentos que arrojen resultados objetivos evaluando las diferentes propuestas utilizadas.

Las métricas utilizadas diferirán en función de la naturaleza del problema a resolver.

En principio, se prevé la utilización en mayor medida de métricas relacionadas con el aprendizaje profundo por refuerzo, pero es posible que exista la necesidad de utilizar otras métricas que permitan la resolución de problemas intermedios e incluso un análisis más minucioso de los resultados obtenidos como velocidad media del vehículo o distancia recorrida.

Entre las posibles métricas de rendimiento de los algoritmos que no utilicen aprendizaje por refuerzo y que serán usados para complementar nuestro sistema podemos encontrar:

* Métricas de clasificación (exactitud, precisión, recuperación, puntuación F1, ROC, AUC,…)
* Métricas de regresión (MSE, MAE)
* Métricas de clasificación (MRR, DCG, NDCG)
* Métricas estadísticas (correlación)
* Métricas de visión por computadora (PSNR, SSIM, IoU)

Entre las posibles métricas de rendimiento específicas de la resolución del problema mediante aprendizaje por refuerzo, serán utilizadas:

* Métricas intrínsecas:
* Recompensa acumulada media durante un episodio
* Sumatorio de la recompensa descontada media durante un episodio
* Tiempo que tarda el algoritmo en converger a una solución óptima durante el entrenamiento
* Estabilidad y fiabilidad del entrenamiento en función del algoritmo y técnicas utilizadas
* Capacidad del agente a generalizar (comportanse bien ante situaciones o escenarios no observados durante el entrenamiento)
* Porcentaje de episodios completados de manera satisfactoria
* Tiempo medio por episodio en alcanzar el objetivo
* Métricas extrínsecas específicas del problema a abordar (para lo cual nos apoyaremos de la herramienta Behavior metrics)
* Solidez del agente ante perturbaciones externas (lluvia, imprevistos, sensores averiados, etc.)
* Distancia recorrida en un circuito concreto
* Velocidad media del vehículo
* Número y gravedad de infracciones cometidas
* Frecuencia de iteración o frames por segundo

Una vez alcanzado un agente óptimo que resuelva el problema en cuestión en un simulador lo más fiel a la realidad posible, se abordará el reto de transferir el conocimiento adquirido a un entorno real con robots que compartan sensores (inputs) y actuadores (outputs) con el entorno simulado durante el entrenamiento.

5. Revistas científicas en las que tiene previsto de forma tentativa publicar los resultados de la tesis doctoral de acuerdo con lo establecido en el artículo 23.3 de la normativa reguladora de los estudios de doctorado de la URJC.

El tipo de investigación desarrollada, teniendo en cuenta su temática, podría publicarse de forma tentativa en las siguientes revistas:

* Neural Computing and Applications, Springer, Q2
* Image and Vision Computing, Elsevier, Q1
* Journal of Artificial Intelligence Research, Q2

6. Plan de trabajo. Describa brevemente las tareas previstas y los resultados esperados en cada bloque temporal.

1º Año:

1. Estudio de estado de la cuestión para el problema del comportamiento automático de robots.
2. Estudio en profundidad de técnicas de deep learning y aprendizaje por refuerzo actuales para tener una base clara sobre la que desarrollar la investigación. Dentro de este estudio incluiremos la realización del curso “machine learning” impartido por Andrew NG en coursera.
3. Implementación y resolución de un problema básico de resolución de un laberinto primero en un simulador propio y luego en Gazebo utilizando técnicas de aprendizaje por refuerzo clásico como Qlearning y Sarsa.
4. Primeros experimentos en Gazebo y RLStudio con circuitos con restricciones y soluciones basadas en reinforcement learning clásico para un robot terrestre móvil.

2º Año:

1. Desarrollo de una herramienta (RL-Studio) que permita la aplicación de algoritmos de reinforcement learning clásico y Deep learning para resolver problemas básicos con los que se pretende recabar técnicas que doten al doctorando de las herramientas necesarias para abordar problemas complejos.
2. Además, este desarrollo dotará a RL-Studio de la potencia necesaria para facilitar la implementación de problemas más complejos que se abordarán en los próximos años. El plan de iteración es el siguiente:

* Generalización de componentes para facilitar la integración de distintos problemas, algoritmos y simuladores de manera rápida y homogénea.
* Implementación de un modo inferencia para ejecutar los algoritmos entrenados sin la sobrecarga relativa a la fase de entrenamiento.
* Entrenamiento y ejecución de agentes en problemas canónicos.

1. Realización del curso “Algorithms, Part 1” impartido por la universidad de Princeton en coursera.

3º Año:

1. Extensión de RL-Studio para probar con distintos simuladores (Gazebo, openAI y CARLA) y algoritmos (entre ellos DQN, DDPG, PPO y A3C).
2. Resolución de problemas básicos propuestos por openAI (mountain car, cartpole y pendulum)
3. Extensión de problemas básicos propuestos por openAI (cartpole y mountain car) para poder extraer de ellos comparativas y lecciones sólidas acerca de las diferentes técnicas y algoritmos a aplicar en problemas más complejos. Entre la extensión de estos problemas se incluirá:

* la posibilidad de moldear los entornos ofrecidos por openAI con el objetivo de complicar el problema básico (e.g configuración el mundo simulado a resolver, mejora de las físicas aplicadas, inclusion un ciclo de reloj en el simulador de openAI que obligue al algoritmo a decidir la acción con rapidez, modificación de la naturaleza del problema de acciones discretas acciones continuas)
* Trasladar el problema a un simulador más realista con físicas más sofisticadas
* Adición de perturbaciones y condiciones adversas

1. Desarrollo de soluciones a problemas más complejos combinando deep learning y reinforcement learning (entre ellos vehículo con autoaparcamiento, sigue-lineas y sigue-carril)
2. Formalización de los problemas resueltos y los bancos de pruebas realizados con objetivo de servir de base para la resolución de futuros problemas y para ilustrar de manera clara el trabajo realizado de manera que se pueda utilizar con objetivos docentes y en publicaciones de diversa índole. Dichos bancos de pruebas consistirán en:

* Aplicación en RL-Studio de distintos algoritmos (solución programática, QLearning, DQN, DDPG, PPO con acciones discretas y PPO con acciones continuas) a la extensíon implementada para el problema cartpole (acciones continuas, físicas mejoradas, aplicación de perturbaciones y condiciones iniciales adversas).
* Aplicación en RL-Studio de distintos algoritmos al problema sigue lineas y sigue carril tanto en Gazebo como en Carla con el objetivo de comparar su rendimiento haciendo uso de las métricas externas evaluadas en Behavior metrics (velocidad efectiva y avance espacial)

1. Realización del curso “Deep reinforcement learning class” impartido por la la compañía Hugging Face.

4º Año:

1. Definición del problema a abordar en la tesis
2. Aplicación de lo previamente aprendido en un simulador con físicas lo más parecido posible al entorno real con el objetivo de facilitar la transferencia de conocimiento al mundo real
3. Publicación de al menos un artículo en el ámbito de la conducción autónoma.

5º Año:

1. Resolución definitiva del problema definido tras haber experimentado durante el año anterior con distintos entornos, técnicas y algoritmos
2. Experimetar sobre la viabilidad de trasladar la política aprendida a un robot móvil en el mundo real.
3. Elaboración de la memoria de la tesis.

7. Bibliografía relevante relacionada con la investigación

Algunas de las referencias más relevantes relacionadas con las investigación incluyen los siguientes trabajos:

* Artificial Intelligence: A Modern Approach by Stuart Russell & Peter Norvig.
* Deep Learning by Ian Goodfellow, Yoshua Bengio and Aaron Courville.
* Reinforcement Learning: An Introduction by Richard S. Sutton and Andrew G. Barto
* End to end Learning for Self-Driving Cars by Mariusz Bojarsky et al.
* Gradient-based learning applied to document recognition by Yann LeCun et al.
* Long Short-Term Memory by Sepp Hochreiter et al.
* DeepPilot: A CNN for Autonomous Drone Racing by Leticia Oyuki Rojas-Perez et al.
* Toromanoff, M., Wirbel, E., & Moutarde, F. (2019). *End-to-End Model-Free Reinforcement Learning for Urban Driving using Implicit Affordances*. http://arxiv.org/abs/1911.10868
* Schulman, J., Moritz, P., Levine, S., Jordan, M. I., & Abbeel, P. (n.d.). *HIGH-DIMENSIONAL CONTINUOUS CONTROL USING GENERALIZED ADVANTAGE ESTIMATION*. Retrieved October 31, 2022, from https://sites.google.com/site/gaepapersupp.
* Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Openai, O. K. (n.d.). *Proximal Policy Optimization Algorithms*.
* Schneider, A., Wickert, C., & Marti, E. (2017). Reducing Complexity by Creating Complexity: A Systems Theory Perspective on How Organizations Respond to Their Environments. *Journal of Management Studies*, *54*(2), 182–208. https://doi.org/10.1111/JOMS.12206
* Rolim, L. A., Santos, F. C. M. dos, Chaves, L. L., Gonçalves, M. L. C. M., Freitas-Neto, J. L., Nascimento, A. L. da S. do, Soares-Sobrinho, J. L., Albuquerque, M. M. de, Lima, M. do C. A. de, & Rolim-Neto, P. J. (2020). Enhanced Reader.pdf. In *Nature* (Vol. 388, pp. 539–547).
* Knox, W. B., Allievi, A., Banzhaf, H., Schmitt, F., & Stone, P. (2021). *Reward (Mis)design for Autonomous Driving*. http://arxiv.org/abs/2104.13906
* Gutiérrez-Moreno, R., Barea, R., López-Guillén, E., Araluce, J., & Bergasa, L. M. (2022). *Reinforcement Learning-Based Autonomous Driving at Intersections in CARLA Simulator*. https://doi.org/10.3390/s22218373
* Engstrom, L., Ilyas, A., Santurkar, S., Tsipras, D., Janoos, F., Rudolph, L., & Adry, A. M. ˛. (n.d.). *IMPLEMENTATION MATTERS IN DEEP POLICY GRADIENTS: A CASE STUDY ON PPO AND TRPO*. Retrieved October 30, 2022, from https://github.com/MadryLab/
* Diederik P. Kingma, J. L. B. (2014). *Adam\_ A Method for Stochastic Optimization \_ Enhanced Reader.pdf*.

Fecha y Firma del doctorando: