Control de drones mediante Reinforcement Learning en plataformas reales

Miguel Fernández Cortizas

Índice general

Resumen

Introducción

gola

Estado del arte

Las teoría clásica de control empleada para estabilizar un cuadricóptero requiere un fino ajuste de los parámetros del modelo. Los últimos avances en aprendizaje automático han permitido que se desarrollen nuevos algoritmos de control empleando técnicas de aprendizaje por refuerzo y redes neuronales.

En 2004, HJ Kim et al. [1] emplearon algoritmos de *reinforcement learing* para estabilizar (*hoover*) un helicóptero y conseguir realizar maniobras acrobáticas.En 2006 Andrew Y. et al [2] siguieron con esta investigación consiguiendo que el helicóptero se estabilizara al revés (*inverted hoover*). En 2010 Travis Dierks et al. [3] desarrollaron un controlador no lineal, basado en redes neuronales, para estabilizar un cuadricóptero y seguir trayectorias.

Unos años después, en 2017 Jemin Hwangbo et al. [4] desarrollaron un método para controlar un quadricóptero con una red neuronal usando técnicas de *reinforcement learning*. En 2018 William Koch et al. [5] desarrollaron un entorno de simulación, GYMFC, para el desarrollo de sistemas de control empleando RL.

Background

3.1. Reinforcement learning

El aprendizaje por refuerzo o *Reinforcement learning* [6] es un área del aprendizaje automático o *Machine Learning* en el que un agente interactúa con un entorno buscando la mejor acción a realizar en función de su estado actual.

Se diferencia de otras técnicas de aprendizaje automático es su enfoque orientado a la interacción directa con el entorno, sin basarse en un modelo completo del entorno o en un conjunto de ejemplos supervisados.

El aprendizaje por refuerzo emplea el marco formal de los procesos de decisión de Markov (*MDP*) en los cuales para definir la interacción entre en agente y el entorno en términos de estados, acciones y recompensas.

Un proceso de decisión de Markov (*MDP*)

Estos procesos de decisión incluyen causalidad, la existencia de recompensas explícitas a sense of uncertanty and nondeterminism

Además del agente y el entorno se pueden identificar cuatro elementos principales más en un sistema de aprendizaje con refuerzo:

- **Política (*Policy*)**. Define el conjunto de acciones que debe realizar el agente para conseguir maximizar su recompensa en función su estado, el cuál es percibido a través del entorno. La *policy* constituye el núcleo del agente y nos permite determinar su comportamiento. Estas políticas pueden ser estocásticas.

- **Recompensa (*Reward signal*)**. Define el objetivo del agente en un problema de aprendizaje por refuerzo. En cada salto de tiempo (*step*) el agente recibe una recompensa por parte del entorno (un número).

  El objetivo del agente es maximizar su recompensa a largo plazo.

- **Función de valor (*Value function*)**. Define el comportamiento que va

- **Policy**. Define el comportamiento que va

Metodología (Problem Formulation)

El objetivo del trabajo es estabilizar un UAV usando algoritmos de control basados en una red neuronal entrenada empleando algoritmos de aprendizaje automático. Los principales componentes que intervienen en el agente son el estado, las acciones y la recompensa, para cada problema hay un conjunto que estados, acciones y funciones de recompensa que pueden llevar a que el agente aprenda.

4.1. Diseño del estado

El UAV cuenta con 2 IMUs para poder obtener datos sobre su estado. Se quiere estabilizar el dron en una orientación concreta, por lo tanto el estado que se ha diseñado consta de 6 parámetros:

$$S = (\varphi, \theta, \psi, \dot{\varphi}, \dot{\theta}, \dot{\psi}) \qquad \varphi, \theta, \psi, \dot{\varphi}, \dot{\theta}, \dot{\psi} \in [-1, 1] \tag{4.1}$$

Siendo φ, θ y ψ los ángulos de alabeo, cabezeo y guiñada del dron y $\dot{\varphi}, \dot{\theta}$ y $\dot{\psi}$ sus respectivas velocidades. Para favorecer la convergencia del aprendizaje, se ha normalizado el estado para que todas sus componentes estén comprendidas dentro del intervalo $[-1, 1]$.

Los ángulos proporcionan información sobre el estado actual y la velocidad angular sobre los estados pasados, es decir, proporciona cierta información temporal.

Para obtener la estimación de orientación se han fusionado las medidas de las 2 IMUs del autopiloto utilizando un filtro complementario, para así conseguir una buena estimación estática junto con una buena respuesta dinámica.

4.2. Diseño de las acciones

Al trabajar con un quadricóptero podemos actuar sobre la potencia que se le entrega a los motores, por lo que cada acción que realice el agente constará de 4 campos:

$$A = (T_1, T_2, T_3, T_4) \qquad T_i \in [-1, 1] \tag{4.2}$$

Siendo T_i la potencia *(Thrust)* normalizada entregada a cada motor. Un valor de $T_1 = -1$ significa que el motor 1 estaría girando a la mínima potencia permitida y un valor de $T_1 = 1$ corresponde a que el motor 1 estaría girando a la máxima potencia.

completar con la transformacion del mundo -1,1 al mundo 1000,2200 μS

4.3. Diseño de la función de recompensa

La función de recompensa rige la forma en la que la red va a configurar sus pesos, por lo tanto, cómo se va a comportar el agente en un estado determinado. Para conseguir que el agente responda de la forma deseada se han probado una gran variedad de funciones de *reward*, optando finalmente por:

$$R_t = \left(1 - \frac{|\varphi| + |\theta| + |\psi|}{3}\right)^3 \tag{4.3}$$

Con esta funcion de recompensa BLA BLA BLA

Hardware

Un cuadricóptero o cuadrirrotor es una aeronave no tripulada (UAV) que está propulsada por cuatro motores cuyas hélices están orientadas verticalmente.Se ha diseñado y construido un cuadrirrotor casero para poder probar en la realidad el control del dron.

Un cuadricóptero convencional cuenta con: un chasis o *frame* que lo sustenta, cuatro motores y la electrónica necesaria para controlarlos, una controladora de vuelo que lo comanda y baterías que le proporcionan energía.

A continuación se detallará como són las distintas partes del dron que se ha fabricado.

5.1. Cuadro

El *frame* está compuesto por perfiles de aluminio y piezas de PLA fabricadas mediante impresión 3D. Cuenta con un nivel para alojar los motores, la receptora de radio y la controladora de vuelo y otro nivel para la batería. Todas las piezas de PLA son de diseño propio y se han realizado mediante software CAD.

imagenes modelo CAD.

5.2. Motores y variadores(ESC)

El dron cuenta con 4 motores sin escobillas (*brushless*) LHI MT2204 II de 2300KV con una tensión de alimentación entre 7.2 V y 11.1 V (2s -3s en una batería LiPo) y una corriente continua máxima de 16A.

Estos motores son trifásicos, es decir, se alimentan con 3 corrientes alternas monofásicas de igual frecuencia y amplitud, desfasadas 120° eléctricos. Para obtener estás formas de ondas a partir de la corriente continua de las baterias, se utilizan los variadores o *ESC*.

(a) Motores LHI MT2204 II empleados

(b) ESC Multistar Race 4 in 1 30A BLHeli empleado

Un variador o *ESC (Electronic speed control)* es un circuito electrónico que controla y regula la velocidad de un motor eléctrico.

Profundicar en las ondas generada por el ESC

5.3. Baterias

Para alimentar al dron, se han elegido baterias tipo LiPo por su alta tasa de descarga (la batería que se ha escogido es capaz de entregar hasta 130 A) y la su estabilidad en la tensión mientras están cargadas.

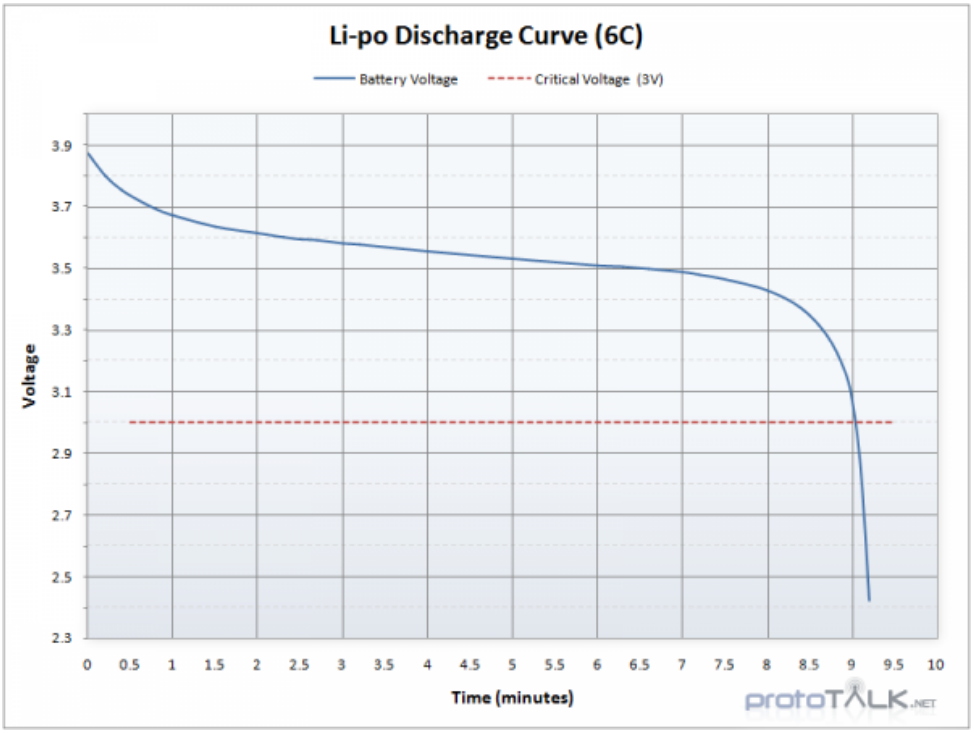

Figura 5.2: Curva típica de descarga de una bateria LiPo (fuente: ProtoTalk.net)

5.4. Autopiloto

En los drones, el sistema que se encarga de estabilizar al cuadricóptero y hacerlo pilotable se denomina la controladora de vuelo o el Autopiloto. Existe una gran variedad de controladoras en el mercado, pero para este trabajo se ha diseñado una controladora propia con el fin de poder tener acceso a todos los sensores y a implementar el algoritmo de control de forma óptima. El autopiloto consta de 3 partes diferenciadas: la electrónica de potencia, el microcontrolador y los sensores. A continuación se tratará sobre estas partes con más detalle.

Estaría bien un par de imágenes de la PCB (anverso y reverso)

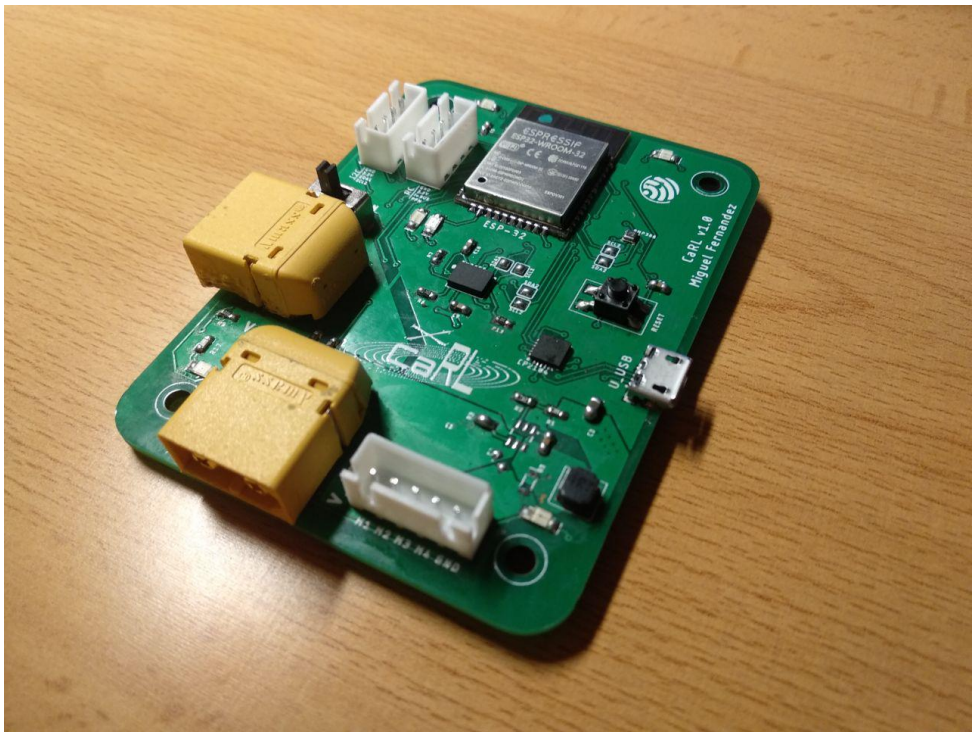

Figura 5.3: Autopiloto CaRL (*Cuadcopter with autopilot based on Reinforcement Learning*).

5.4.1. Fase de Potencia

Con el fin de poder gestionar la potencia entregada por las baterias a la placa y a los motores se ha diseñado una etapa de potencia en la que se debe mencionar dos partes: el interruptor de potencia y el regulador a 3.3 Voltios.

Interruptor de potencia

Los motores del dron pueden llegar a consumir 12 Amperios cada uno, lo que los cuatro motores pueden llegar a consumir 48 Amperios. Un interruptor con tamaño reducido no puede manejar tanta corriente, por ello se ha empleado un transistor MOSFET de canal P por el que pueden circular hasta 100 Amperios, con el fin de abrir o cerrar el

paso de corriente desde las baterías al resto de la placa. El MOSFET se controla con un interruptor de poca potencia entre drenador y puerta.

Cuando se cierra el interruptor se alimenta directamente al ESC y al regulador de tensión.

Regulador a 3.3V

La eléctronica digital de la PCB se alimenta y emplea lógica a 3.3 Voltios, por lo que no la podemos conectar a las baterías de 11.1 Voltios. Para adecuar la tensión se ha escogido un regulador Step-down de tipo Buck (Figura 5.4) (¿explico como funciona un convertidor Buck?). El circuito integrado que se encarga de conmutar la fuente es el chip AP3211.

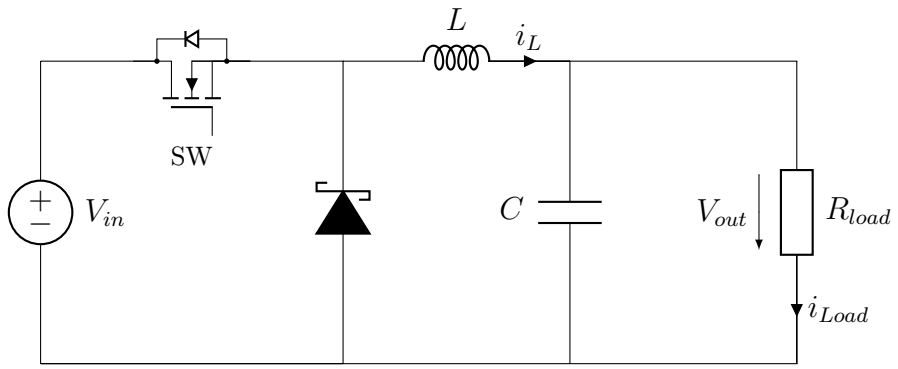

Figura 5.4: Esquema de un convertidor Buck

5.4.2. El microcontrolador (ESP32)

El microcontrolador por el que se ha optado para este Autopiloto es el ESP32, un microcontrolador de doble núcleo con dos CPUs XTensaL6 con arquitectura Harvard [7]. El ESP32 tiene una frecuencia de reloj de hasta 240MHz ,y cuenta con una antena WiFi a 2,4 GHz y conexión Bluetooth 4.2 BLE [8]. Los motivos por los que se ha decidido emplear este microcontrolador son:

- Elevada frecuencia de procesamiento y dos nucleos de procesamiento.
- Antena WiFi incorporada.
- Bajo consumo de potencia.

Para poder programar el microcontrolador se utiliza un convertidor USB (Bus Serie Universal) a UART (Transmisor-Receptor Asíncrono Universal) que permite conectar por USB el microcontrolador para poder programarlo y hacer depuración utilizando comunicaciones Serial. El chip que realiza esta funcion es el CP2104.

5.4.3. Sensores

La principal fuente de información procedente del exterior que recibe una controladora de vuelo se la proporcionan las unidades de medición inercial (IMU). Las IMUs son

dispositivos electrónicos que son capaces de medir aceleraciones, velocidades y detectar la orientación de un sistema. El principal problema de estos sensores es que suelen sufrir error acumulativo. ¿profundizo en los sensores MEMS (imus electrónicas)?

Otros sensores utilizados frecuentemente en los autopilotos son brújulas (se encuentran integrados en la IMU para corregir errores de orientación) y barómetros (para estimar la altitud a la que se encuentra el dron).

Nuestro autopiloto cuenta con dos IMUs de 9 Grados de Libertad y un barómetro para conseguir una mejor estimación del estado del cuadricóptero:

1. **BNO 055 (BOSCH)**: El circuito integrado de Bosch es un sensor ïnteligente"que incluye los sensores y la fusión de las lecturas de los distintos sensores en un único componente. Este sensor nos proporciona estimaciones del estado con muy poca deriva. argumentar un poco mejor
2. **MPU 9250 (TDK InvenSense)**: El sensor inercial de TDK tiene una mejor respuesta dinámica, aunque la fusión de las lecturas del sensor se realiza externamente en el microcontrolador del dron.
3. **BMP388 (BOSCH)**:

5.5. Banco de pruebas

Para poder realizar la experimentación real de forma segura, se ha diseñado una base para sujetar al drone permitiendo que rote en roll pitch y yaw de forma restringida.

La estructura de sujección esta formado por 2 juntas esféricas acopladas una a continuación de la otra, para permitir la rotación en el espacio.

imagen rotulas y/o CAD

Los límites físicos de las rótulas permiten que el dron tome orientaciones comprendidas entre:

$$\begin{aligned} -60^\circ \leq \varphi \leq 60^\circ \quad & \text{Roll} \\ -60^\circ \leq \theta \leq 60^\circ \quad & \text{Pitch} \\ -180^\circ \leq \psi \leq 180^\circ \quad & \text{Yaw} \end{aligned}$$

Experimentos

Pa

Discusión

Conclusiones y trabajo futuro

Bibliografía

[1] H. J. Kim, M. I. Jordan, S. Sastry, and A. Y. Ng, "Autonomous helicopter flight via reinforcement learning," in *Advances in neural information processing systems*, 2004, pp. 799–806.

[2] A. Y. Ng, A. Coates, M. Diel, V. Ganapathi, J. Schulte, B. Tse, E. Berger, and E. Liang, "Autonomous inverted helicopter flight via reinforcement learning," in *Experimental robotics IX*. Springer, 2006, pp. 363–372.

[3] T. Dierks and S. Jagannathan, "Output feedback control of a quadrotor uav using neural networks," *IEEE transactions on neural networks*, vol. 21, no. 1, pp. 50–66, 2010.

[4] J. Hwangbo, I. Sa, R. Siegwart, and M. Hutter, "Control of a quadrotor with reinforcement learning," *IEEE Robotics and Automation Letters*, vol. 2, no. 4, pp. 2096–2103, 2017.

[5] W. Koch, R. Mancuso, R. West, and A. Bestavros, "Reinforcement learning for uav attitude control," *ACM Transactions on Cyber-Physical Systems*, vol. 3, no. 2, p. 22, 2019.

[6] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.

[7] "Esp32 technical reference manual," Espressif Systems, 2018. [Online]. Available: https://www.espressif.com/sites/default/files/documentation/esp32_technical_reference_manual_en.pdf

[8] "Esp32 datasheet," Espressif Systems, 2019. [Online]. Available: https://www.espressif.com/sites/default/files/documentation/esp32_datasheet_en.pdf