Técnicas de Control por Aprendizaje por Refuerzo y Programación Dinámica

Programa de Doctorado en Automática, Robótica e Informática Industrial

Doctorando: **Henry Díaz I.** Directores: Antonio Sala[1], Leopoldo Armesto[2]
email:hendia@posgrado.upv.es

[1]Instituto de Automática e Informática Industrial(AI[2]), [2]Instituto de Diseño y Fabricación(IDF)

Introducción

La programación dinámica (DP, *dynamic programming*) y el aprendizaje por refuerzo (RL, *reinforcement learning*) son un conjunto de técnicas para resolver problemas de decisión secuenciales.
Este tipo de problemas secuenciales aparecen en una amplia variedad de campos entre los que podemos mencionar el control automático, control teórico, inteligencia artificial, robótica, entre otras.

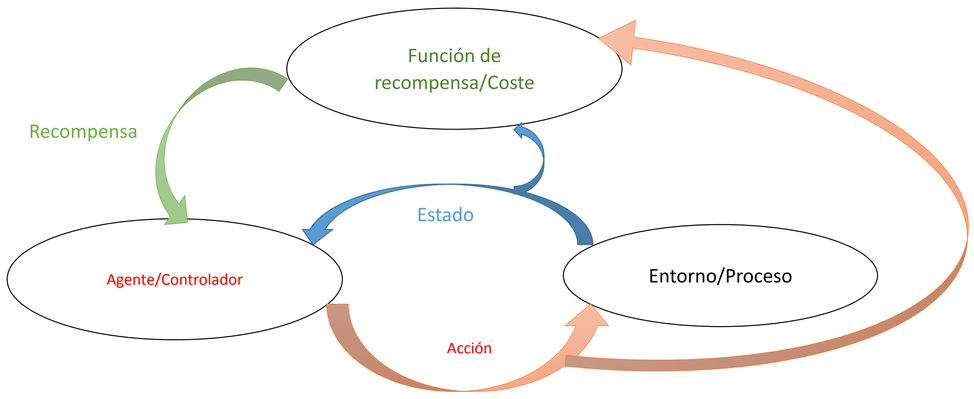

Figure 1: Elementos de la DP y RL y su flujo de interacción.

Objetivos

El objetivo de la investigación es desarrollar nuevas técnicas de control óptimo no lineal mediante RL y DP aplicada a robots y/o sistemas electromecánicos.
Se desarrollará y analizará algoritmos y su convergencia, análisis de estabilidad en sistemas no lineales, aprendizaje offline y online, algoritmos con horizonte de tiempo finito que de una u otra manera tiene una conexión con el control dinámico adaptativo.

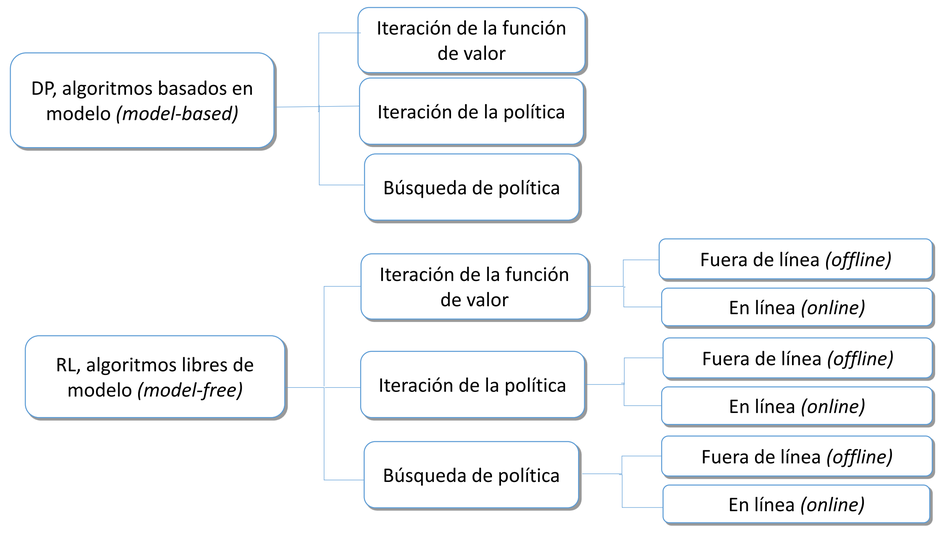

Figure 3: Taxonomía de los algoritmos de DP y RL.

Plataforma Experimental

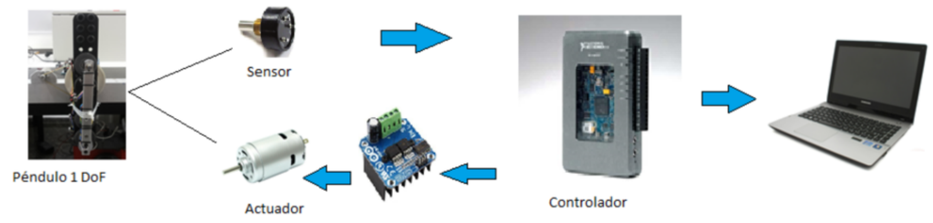

Figure 2: Sistema Electromecánico. Péndulo invertido de 1DoF.

Resultados Experimentales

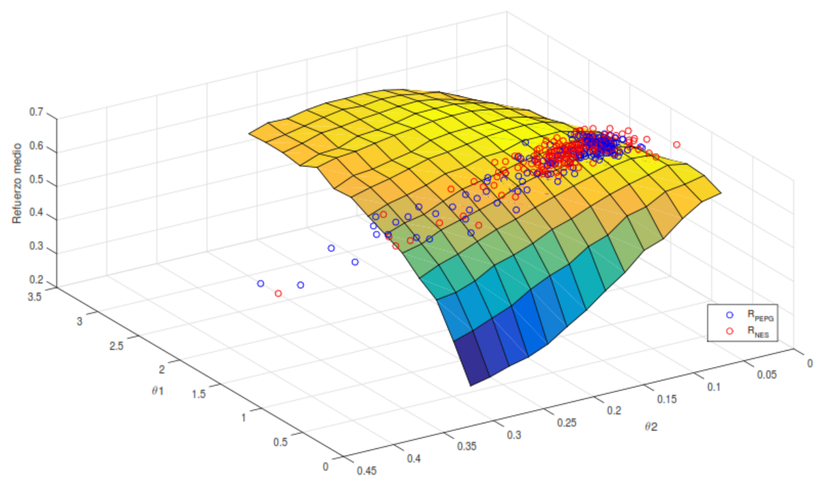

Figure 4: Proceso de aprendizaje de las ganancias de un controlador aplicando técnicas de búsqueda directa de políticas a un péndulo de 1DoF.

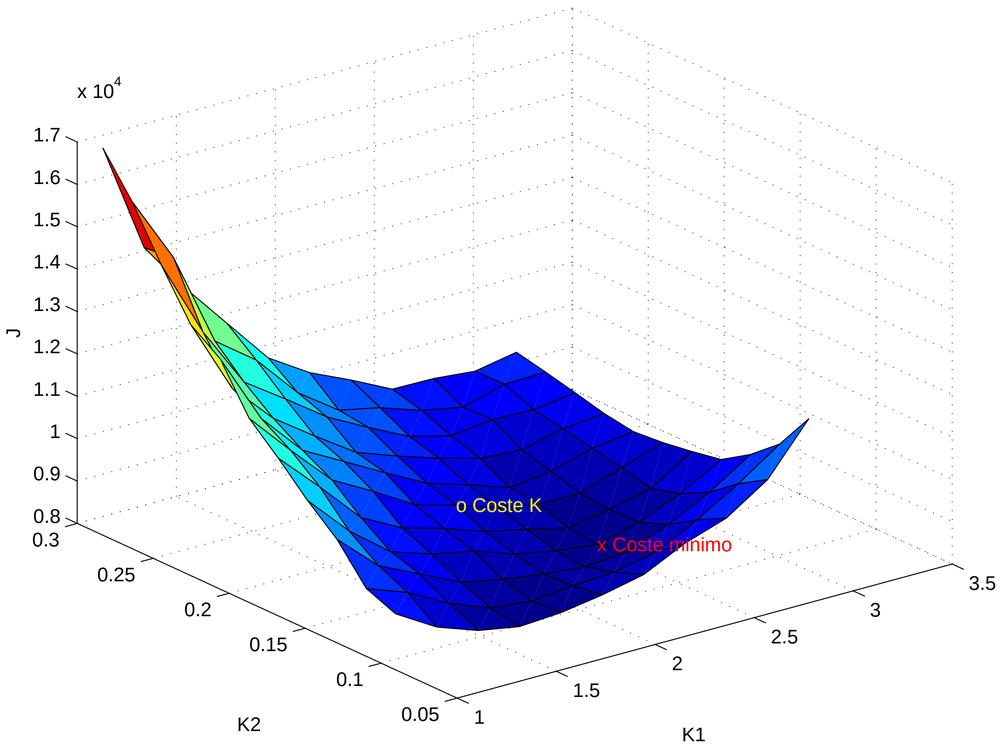

Figure 5: Superficie de costes obtenidos al variar las ganancias del controlador entorno a las ganancias aprendidas por el algoritmo Q-Learning (técnica de aproximación de la función de valor) en el péndulo de 1 DoF.

Conclusiones

- El Aprendizaje por Refuerzo no requiere de un modelo de comportamiento del sistema, sino que funcionan únicamente empleando datos obtenidos del entorno, a diferencia de las técnicas de programación dinámica las cuales proporcionan soluciones basadas en el modelo de comportamiento del sistema.
- Cuando el número de estados que posee un problema es muy grande o infinito, no resulta viable almacenar las funciones de valor y las políticas con tablas, por lo que surge la necesidad del uso de aproximadores funcionales con las ventajas y desventajas que su uso implica.

Referencias

[1] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press Cambridge, 1998, vol. 1, no. 1.

[2] D. P. Bertsekas and Bertsekas, *Dynamic programming and optimal control*. Athena Scientific Belmont, MA, 1995, vol. 1, no. 2.

[3] F. L. Lewis and D. Liu, *Reinforcement learning and approximate dynamic programming for feedback control*. John Wiley & Sons, 2013, vol. 17.

[4] L. Busoniu, R. Babuska, B. De Schutter, and D. Ernst, *Reinforcement learning and dynamic programming using function approximators*. CRC press, 2010, vol. 39.

[5] M. P. Deisenroth, G. Neumann, J. Peters *et al.*, "A survey on policy search for robotics." *Foundations and Trends in Robotics*, vol. 2, no. 1-2, pp. 1–142, 2013.

Agradecimientos

Los autores agradecen al Gobierno de España, MINECO(DPI2016-81002-R) y al Gobierno de Ecuador(Beca SENESCYT) la financiación recibida para llevar a cabo la investigación.