

Tecnológico de Costa Rica
Escuela de Ingeniería Electrónica
Programa de Licenciatura en Ingeniería Electrónica



**Método basado en aprendizaje reforzado
para el control automático de una
planta no lineal**

Informe de Trabajo Final de Graduación para optar por el título de
Ingeniero en Electrónica con el grado académico de Licenciatura

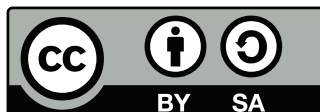
Oscar Andrés Rojas Fonseca

Borrador de 15 de mayo de 2024

El documento **Requisitos para la entrega de Trabajos Finales de Graduación** a las bibliotecas del TEC indica que usted debe incluir la licencia de Creative Commons en la página siguiente de la portada.

Asegúrese entonces de **elegir la licencia correcta**, y ajustar el texto abajo a su selección.

Es necesario que **descargue el ícono** correcto en formato vectorial, y lo coloque en el directorio **fig/**.



Este trabajo titulado *Método basado en aprendizaje reforzado para el control automático de una planta no lineal* por Oscar Andrés Rojas Fonseca, se encuentra bajo la Licencia Creative Commons **Atribución-ShareAlike 4.0 International**.

Para ver una copia de esta Licencia, visite <http://creativecommons.org/licenses/by-sa/4.0/>.

Declaro que el presente documento de tesis ha sido realizado enteramente por mi persona, utilizando y aplicando literatura referente al tema e introduciendo conocimientos y resultados experimentales propios.

En los casos en que he utilizado bibliografía he procedido a indicar las fuentes mediante las respectivas citas bibliográficas. En consecuencia, asumo la responsabilidad total por el trabajo de tesis realizado y por el contenido del presente documento.

Oscar Andrés Rojas Fonseca

Cartago, 15 de mayo de 2024

Céd: 1-1696-0962

Instituto Tecnológico de Costa Rica
Escuela de Ingeniería Electrónica
Trabajo Final de Graduación
Acta de Aprobación

Defensa de Trabajo Final de Graduación
Requisito para optar por el título de Ingeniero en Electrónica
Grado Académico de Licenciatura

El Tribunal Evaluador aprueba la defensa del trabajo final de graduación denominado *Método basado en aprendizaje reforzado para el control automático de una planta no lineal*, realizado por el señor Oscar Andrés Rojas Fonseca y, hace constar que cumple con las normas establecidas por la Escuela de Ingeniería Electrónica del Instituto Tecnológico de Costa Rica.

Miembros del Tribunal Evaluador

Dra. María Curie Pérez
Profesora Lectora

M.Sc. Pedro Pérez Pereira
Profesor Lector

Ing. Albert Einstein Sánchez
Profesor Asesor

Cartago, 15 de mayo de 2024

Instituto Tecnológico de Costa Rica
Escuela de Ingeniería Electrónica
Trabajo Final de Graduación
Tribunal Evaluador
Acta de Evaluación

Defensa del Trabajo Final de Graduación
Requisito para optar por el título de Ingeniero en Electrónica
Grado Académico de Licenciatura

Estudiante: **Oscar Andrés Rojas Fonseca** Carné: 2018102187

Nombre del proyecto: *Método basado en aprendizaje reforzado para el control automático de una planta no lineal*

Los miembros de este Tribunal hacen constar que este trabajo final de graduación ha sido aprobado y cumple con las normas establecidas por la Escuela de Ingeniería Electrónica del Instituto Tecnológico de Costa Rica y es merecedor de la siguiente calificación:

Nota del Trabajo Final de Graduación: _____

Miembros del Tribunal Evaluador

Dra. María Curie Pérez
Profesora Lectora

M.Sc. Pedro Pérez Pereira
Profesor Lector

Ing. Albert Einstein Sánchez
Profesor Asesor

Cartago, 15 de mayo de 2024

Resumen

El resumen es la síntesis de lo que aparece en el resto del documento. Tiene que ser lo suficientemente conciso y claro para que alguien que lo lea sepa qué esperar del resto del trabajo, y se motive para leerla completamente. Usualmente resume lo más relevante de la introducción y contiene la conclusión más importante del trabajo.

Es usual agregar palabras clave, que son los temas principales tratados en el documento. El resumen queda fuera de la numeración del resto de secciones.

Evite utilizar referencias bibliográficas, tablas, o figuras en el resumen.

Palabras clave: palabras, clave, energía, cambio climático, RISC V

Abstract

Same content as the Spanish version, just in English. Check [this site](#) for some help with the translation. For instance, the following is the automatic translation from a previous version of the “Resumen”.

The abstract is the synthesis of what appears in the rest of the document. It has to be concise and clear enough so that someone reading it knows what to expect from the rest of the text, and is motivated to read it in full. It usually summarizes the most relevant parts of the introduction and contains the most important conclusion of the work.

It is usual to add keywords, which are the main topics covered in the document. The abstract is left out of the numbering of the rest of the sections.

Avoid using bibliographical references, tables, or figures in the abstract.

Keywords: word 1, word 2,

a mis queridos padres

Agradecimientos

El resultado de este trabajo no hubiese sido posible sin el apoyo de Thevenin, Norton, Einstein y mi querido amigo Ohm.

Usualmente se agradece aquí a la empresa o investigador que dio la oportunidad de realizar el trabajo final de graduación.

No debe confundir el agradecimiento con la dedicatoria. La dedicatoria es usualmente una sola línea, con la persona a quien se dedica el trabajo.

El agradecimiento es un texto más elaborado, de carácter personal, en donde se expresa la gratitud por la oportunidad, el apoyo brindado, la inspiración ofrecida, el acompañamiento moral, etc.

Oscar Andrés Rojas Fonseca

Cartago, 15 de mayo de 2024

Índice general

Índice de figuras	II
Índice de tablas	III
Revisar	IV
1. Introducción	2
1.1. *****	2
2. Marco teórico	3
2.1. Péndulo amortiguado a hélice PAMH	3
2.2. Aprendizaje reforzado RL	4
2.3. *****	5
2.3.1. DQN	5
2.3.2. PPO	6
3. Solución propuesta	7
4. Resultados y análisis	8
5. Conclusiones	9
Bibliografía	10
A. Demostración del teorema de Nyquist	11

Índice de figuras

2.1. Modelo simplificado del PAMH.	3
2.2. Resumen de categorización del RL DataScience	4
2.3. Esquemático del control predictivo del modelo (MPC) DataScience	5

Índice de tablas

Revisar

Alfonso Chacón

Definición de problema a resolver y enfoque de solución Problema un poco abierto quizás pero claro. A como está planteado, puede disputarse si la elección de un método de aprendizaje reforzado es la óptima (algo igual no tan relevante). ¿Qué se gana con usar un método como los propuestos en un problema ya resuelto por métodos clásicos? ¿Más allá de la obvia consideración didáctica? (o simplemente: lo hacemos porque lo podemos hacer). Considero conveniente contestar esta pregunta de manera explícita en el desarrollo del proyecto.

Objetivos y otros Objetivo general: Adecuado. Se puede mejorar dándole un poco más de especificidad, es decir, aclarar que el método a usar es DRL. Pero nada serio.

Redacción Buena. Algunos detalles de titulación mínimos (en español, solo se capitulan la primera letra del título y los nombres propios en los mismos).

Criterio: Aprobado 27/02/2024

Javier Pérez

Entorno del problema El Entorno del problema es muy largo, hay mucho texto que no es necesario para entender las circunstancias del proyecto. Sea más directo.

Planteamiento del problema Los dos primeros párrafos del Planteamiento del problema es más parte de un marco teórico.

Soluciones La redacción de las posibles soluciones y Selección de la solución se vuelve confusa por momentos. El diagrama que pone de la solución es más bien un esquema de la secuencia de pasos que se realizarán. No es lo mismo.

Criterio: Aprobado 05/02/2024

Capítulo 1

Introducción

Bajo dicha modalidad de trabajo se espera el cumplimiento de algunos requisitos necesarios para un correcto entrenamiento y validación del modelo a elaborar, los cuales se muestran en el cuadro ??

1.1. *****

Capítulo 2

Marco teórico

2.1. Péndulo amortiguado a hélice PAMH

El péndulo amortiguado a hélice corresponde a una planta de laboratorio compuesta de un motor con hélice controlado por torque, una masa pequeña, péndulo y soportes de aluminio de baja fricción. Un modelo simplificado del sistema se muestra en la Figura 2.1 PAMH1.

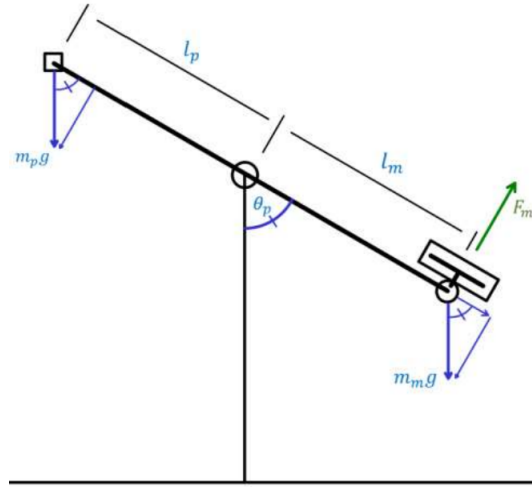


Figura 2.1: Modelo simplificado del PAMH.

El objetivo de dicho sistema es controlar la magnitud del ángulo θ_p , únicamente ejerciendo torque al accionar a una distancia l_m el motor con una fuerza F_m y movimiento de su masa m_m , mientras a una distancia de l_p del centro se encuentra una masa m_p que contrarresta el movimiento.

De manera que al analizar el sistema con sumatoria de torques se obtiene la constante de rosamiento central B (en caso de existir) junto con la inercia ejercida J_p . Por lo tanto, se definen las variables de estado siguientes y sus ecuaciones de estado mostradas en 2.1 ControlModerno.

$$\begin{aligned}
x_1 &= \theta_p & x_2 &= \dot{\theta}_p & y &= x_1 = \theta_p \\
\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = -\frac{B}{J_p}x_2 + (m_p l_p - m_m l_m) \frac{g}{J_p} \sin(x_1) + \frac{l_m}{J_p} F_m \end{cases}
\end{aligned} \tag{2.1}$$

2.2. Aprendizaje reforzado RL

Al estudiar el concepto de aprendizaje reforzado y los diferentes métodos y algoritmos que corresponden a este tipo de aprendizaje automático, se obtiene el resumen de la Figura 2.2, en donde se muestra que las principales secciones son el RL basado en modelo y el libre de modelo. De igual forma se cuenta con el aprendizaje reforzado profundo (DRL), una combinación y reestructuración de métodos de cada subdivisión [DataScience](#).

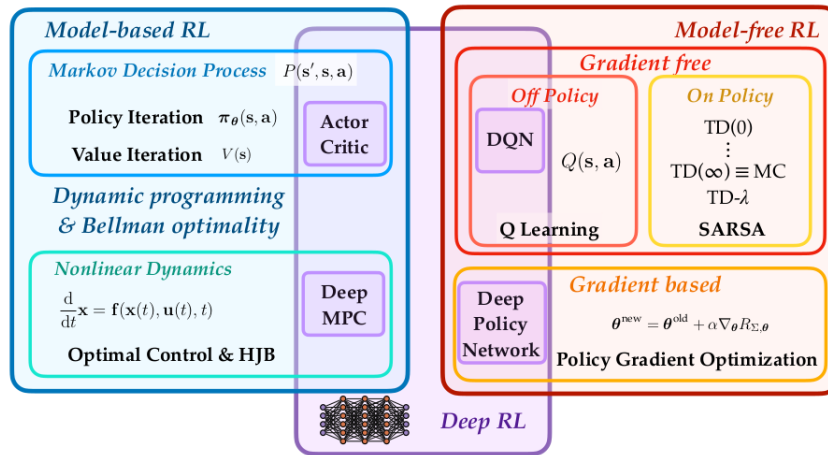


Figura 2.2: Resumen de categorización del RL [DataScience](#).

De igual manera, los avances en la investigación de diferentes métodos como las redes neuronales recurrentes (RNN), ejemplificado por Mamba [Mamba](#), ha mostrado la capacidad de optimización del desempeño de estas para llegar a competir con los modelos basados en *Transformer*.

Control predictivo del modelo (MPC)

La estructura de comunicación entre el sistema y el modelo de control mediante maching learning se efectua con base en funciones de optimización, donde se generan acciones controladas en cada paso para controlar alguna característica ($\hat{\mathbf{F}}$) y se toma la primera muestra de dicha acción (\hat{x}_{k+1}) para variar la respuesta del controlador enviada al sistema (\mathbf{u}_j), esto independiente y aplicable a sistemas lineales o no lineales. El proceso se simplifica en la Fig. 2.3.

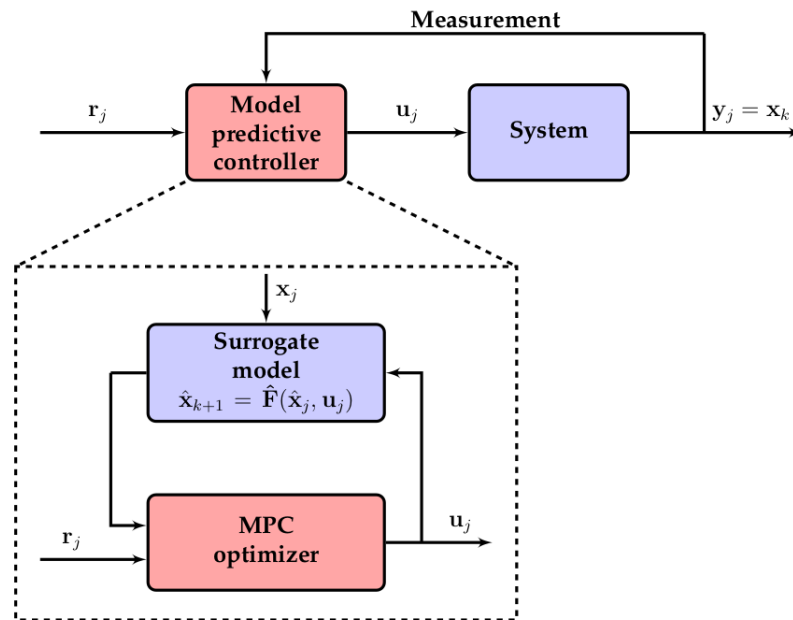


Figura 2.3: Esquemático del control predictivo del modelo (MPC) [DataScience](#).

2.3. *****

2.3.1. DQN

DQN es un algoritmo de aprendizaje por refuerzo basado en valores introducido por DeepMind en 2013. Se diseñó específicamente para aprender a jugar a juegos de Atari directamente a partir de entradas de píxeles sin procesar³. Estos son los aspectos clave de DQN:

Q-Learning y Diferencia Temporal: La "Q" en DQN significa "Q-Learning", que es un método de diferencia temporal fuera de política. DQN tiene en cuenta las recompensas futuras al actualizar la función de valor para un par estado-acción determinado. A diferencia de los métodos de gradiente de política (por ejemplo, REINFORCE), DQN no requiere esperar hasta el final de un episodio para calcular la recompensa final. **Atari Games y Raw Pixels:** DQN ganó popularidad al aprender con éxito a jugar a juegos de Atari utilizando datos de píxeles en bruto como entrada. El agente aprende a realizar acciones basándose en los valores de píxeles observados, sin ninguna característica creada a mano. La ecuación de Bellman guía las actualizaciones de la función de valor a medida que el agente interactúa con el entorno. **Ejemplo de MountainCar-v0:** Consideremos el entorno MountainCar-v0 de OpenAI Gym. En esta tarea, un coche con poca potencia debe subir una colina empinada cogiendo suficiente impulso. El agente recibe una recompensa de -1 por cada acción realizada hasta que alcanza la bandera (recompensa 0). El episodio termina si el agente no consigue alcanzar la montaña en 200 pasos.

2.3.2. PPO

Función objetivo: PPO utiliza una función objetivo sustituta para restringir el tamaño del paso durante las actualizaciones de la política. Esto garantiza un aprendizaje estable y eficaz. A diferencia de otros algoritmos, PPO no requiere esperar hasta el final de un episodio para calcular la recompensa final, sino que actualiza la función de valor de forma incremental a medida que el agente interactúa con el entorno. Estimación de ventaja generalizada (GAE): Aunque el documento original de la PPO hace abstracción de la estimación de ventajas en su objetivo, la aplicación real utiliza GAE. GAE ayuda a estimar las ventajas de diferentes acciones basándose en las recompensas y funciones de valor observadas. Normalización de ventajas: Después de calcular las ventajas utilizando GAE, PPO normaliza las ventajas restando la media y dividiendo por la desviación estándar. Cabe destacar que esta normalización se produce a nivel de minilotes, no a nivel de lotes completos².

AAAA [1]

Capítulo 3

Solución propuesta

Primero que todo, jamás utilice el título indicado arriba, sino algo relacionado con su solución: “Sistema de corrección de distorsión” o lo que competa a su tesis en particular.

Este capítulo puede separarse en varias secciones, dependiendo del problema concreto. Aquí los algoritmos o el diseño del sistema deben quedar lo suficientemente claros para que otra persona pueda re-implementar al sistema propuesto. Sin embargo, el enfoque no debe nunca concentrarse en los detalles de la implementación particular realizada, sino del diseño conceptual como tal.

Recuerdese que toda tabla y figura debe estar referenciada en el texto [2].

Capítulo 4

Resultados y análisis

En este capítulo se exponen los diseños experimentales realizados para comprobar el funcionamiento correcto del sistema. Por ejemplo, si se realiza algún sistema con reconocimiento de patrones, usualmente esta sección involucra las llamadas *matrices de confusión* donde se compactan las estadísticas de reconocimiento alcanzadas. En circuitos de hardware, experimentos para determinar variaciones contra ruido, etc. También pueden ilustrarse algunos resultados concretos como ejemplo del funcionamiento de los algoritmos. Puede mostrar por medio de experimentos ventajas, desventajas, desempeño de su algoritmo, o comparaciones con otros algoritmos.

Recuerde que debe minimizar los “saltos” que el lector deba hacer en su documento. Por tanto, usualmente el análisis se coloca junto a tablas y figuras presentadas, y debe tener un orden de tal modo que se observe cómo los objetivos específicos y el objetivo general del proyecto de tesis se han cumplido.

Capítulo 5

Conclusiones

Las conclusiones no son un resumen de lo realizado sino a lo que ha llevado el desarrollo de la tesis, no perdiendo de vista los objetivos planteados desde el principio y los resultados obtenidos. En otras palabras, qué se concluye o a qué se ha llegado después de realizado la tesis de maestría. Un error común es “concluir” aspectos que no se desarrollaron en la tesis, como observaciones o afirmaciones derivadas de la teoría directamente. Esto último debe evitarse.

Es fundamental en este capítulo hacer énfasis y puntualizar los aportes específicos del trabajo.

Es usual concluir con lo que queda por hacer, o sugerencias para mejorar los resultados.

Bibliografía

- [1] J. Schulman, F. Wolski, P. Dhariwal, A. Radford y O. Klimov, *Proximal Policy Optimization Algorithms*, 2017. arXiv: [1707.06347](#) [cs.LG].
- [2] C. S. Burrus, J. H. McClellan, A. V. Oppenheim, T. W. Parks, R. W. Schafer y H. W. Schuessler, *Ejercicios de Tratamiento de la Señal. Un enfoque práctico*. Prentice Hall, 1998.

Apéndice A

Demostración del teorema de Nyquist

El título anterior es solo un ejemplo ilustrativo. Éste teorema no ameritaría un apéndice pues es parte normal del currículum de Electrónica, pero apéndices usualmente involucran aspectos de esta índole, que se salen de la línea de la tesis, pero que es conveniente incluir por completitud.

Los anexos contienen toda información adicional que se considere pertinente agregar, como manuales de usuario, demostraciones matemáticas que se salen de la línea principal de la tesis, pero que pueden considerarse parte de los resultados del trabajo.