

Tabla de Contenido

1. Introducción	1
1.1. Motivación	3
1.2. Definición del escenario a estudiar	3
1.3. Hipótesis de investigación	4
1.4. Objetivo general	4
1.5. Objetivo específicos	4
1.6. Metodología y herramientas	5
1.6.1. <i>OpenAI GYM</i>	5
1.6.2. <i>Pytorch</i>	5
2. Marco teórico y estado del arte	7
2.1. Modelo Fenomenológico del estanque	7
2.1.1. Constantes del sistema	8
2.2. Control PID	8
2.3. Aprendizaje reforzado	9
2.3.1. Proceso de decisión Markoviano	9
2.3.2. Loop de aprendizaje	10
2.3.3. Exploración ϵ -greedy	11
2.3.4. Algoritmos populares	11
2.3.4.1. Gradiente de la política	11
2.3.4.2. Aproximación mediante programación dinámica	12
2.3.4.3. Algoritmos de actor-critico	13
2.3.4.4. Aprendizaje reforzado basado en modelos	13
2.4. Aprendizaje Reforzado fuera de la política	13
2.4.1. Deep Q-Network	13
2.5. Aprendizaje Reforzado fuera de línea	14
2.5.1. Incapacidad de explorar	14
2.5.2. <i>Distributional shift</i>	14
2.5.3. <i>Random Ensemble Mixture</i>	14
2.6. Estado del arte	15
2.6.1. Métodos de restricción	15
2.6.1.1. Batch Constrained deep Q-learning (BCQ)	15
2.6.1.2. Bootstrapping Error Accumulation Reduction (BEAR)	16
2.6.2. Métodos enfocados en datos	16
2.6.2.1. Scaling Data-driven Robotics with Reward Sketching and Batch Reinforcement Learning	16

2.6.2.2.	REM-DQN	17
2.6.3.	Futuro y benchmarks	17
2.7.	Uso del marco teórico	18
3.	Metodología	19
3.1.	Diseño del experimento	19
3.1.1.	Creación del Ambiente virtual	19
3.1.2.	Generación de datos	20
3.1.3.	Estanque DQN dataset	21
3.1.4.	Estanque PIDs dataset	21
3.1.5.	Entrenamiento offline	22
3.1.6.	Inferencia	23
3.2.	Metodología de análisis	23
3.2.1.	Desempeño	23
3.2.2.	Comportamiento frente a eventualidades	24
4.	Resultados y análisis	25
4.1.	Creación de Datos mediante entrenamiento de un agente online	25
4.1.1.	Funciones de mérito	25
4.1.2.	DQN FPID	26
4.1.3.	DQN CFR	27
4.2.	Creación de Datos mediante PID	28
4.3.	Entrenamiento REM-DQN offline	29
4.3.1.	REM-DQN usando datos DQN-CFR	29
4.3.2.	REM-DQN usando datos DQN-FPID	30
4.3.3.	REM-DQN usando datos PID	31
4.4.	Comparación Control	31
4.4.1.	Desempeño	31
4.4.2.	Comportamiento frente a eventualidades	33
4.4.2.1.	Cambio en β	33
4.4.2.2.	Cambio en c_1	34
4.4.2.3.	Cambio en c_2	35
4.4.2.4.	Cambio en α	36
4.4.3.	Análisis del desempeño	37
5.	Conclusiones	40
	Bibliografía	42