FutuRed

III EDICIÓN CONGRESO DE

REDES INTELIGENTES

III WEBINAR 11 DE JUNIO 2025 - FUTURED

Congreso organizado por: ITE CENTRO TECNOLÓGICO

Patrocinado por: Schneider Electric · ormazabal velatia · gridspertise · SIEMENS

Automatizando la operación de la red de transporte de electricidad

- Ponente(s) / organización(es)
  - Eloy Anguiano Batanero / IIC & UAM

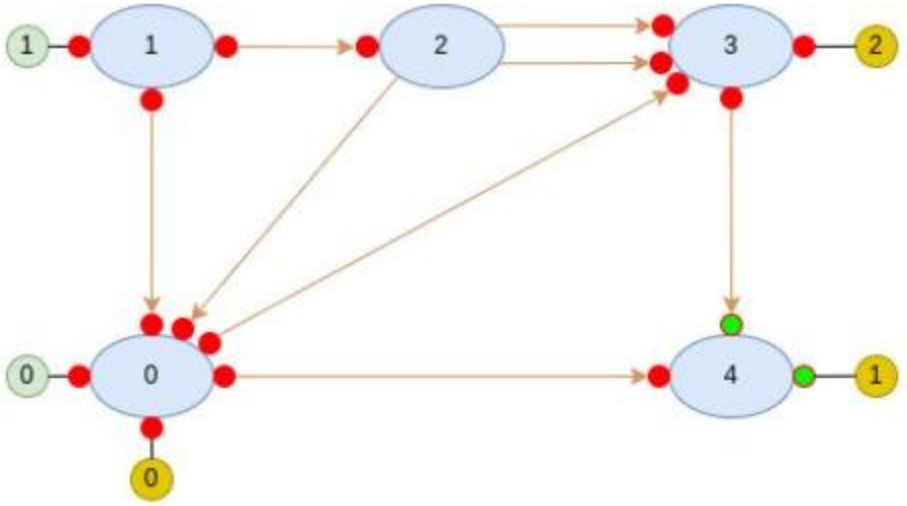
- Autor(es) / organización(es)
  - Eloy Anguiano Batanero / IIC & UAM
  - Ángela Fernández Pascual / UAM
  - Álvaro Barbero Jiménez / IIC

instituto de ingeniería del conocimiento

Universidad Autónoma de Madrid

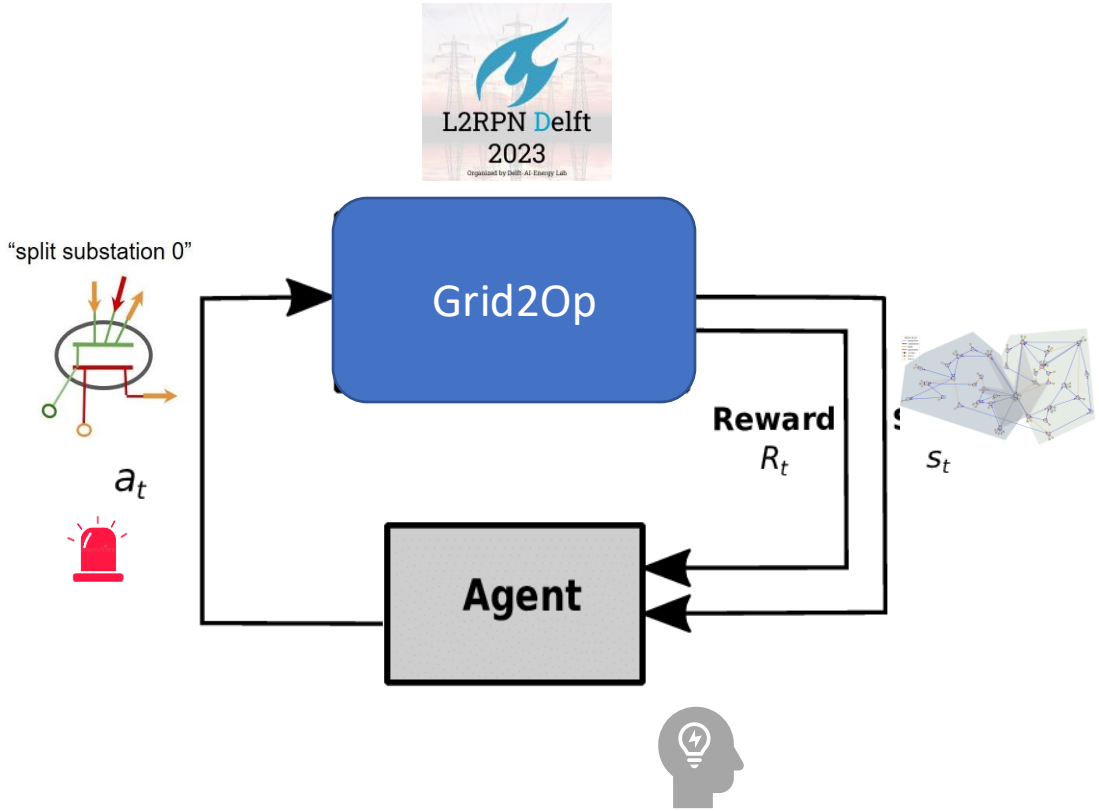
1. Objetivo

- Motivación:
  - Creciente penetración de renovables – Mayor volatilidad y complejidad operativa.
  - Necesidad de herramientas que ayuden a la operativa en tiempo real.
- Propósito de estudio:
  - Probar la utilidad de un agente de aprendizaje por refuerzo para aprender políticas:
    - Robustas ante contingencias no vistas.
    - Promuevan la reducción de las pérdidas eléctricas.
- Caso de estudio:
  - Entorno Grid2Op – rte_case5_example
    - 5 subestaciones.
    - 13 elementos.
    - 20 crónicas de 2016 pasos de 5 minutos.



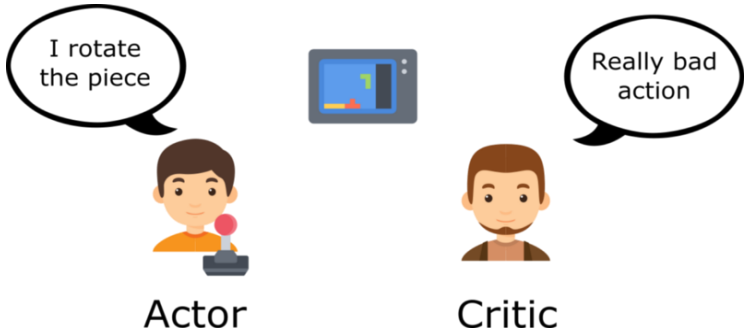
2. Metodología: Aprendizaje por refuerzo

- Proceso de decisión de Markov (MDP):
  - Proceso iterativo compuesto por:
    - Estado (S)
    - Acción (A)
    - Recompensa (R)
  - Permite al agente descubrir su propia estrategia



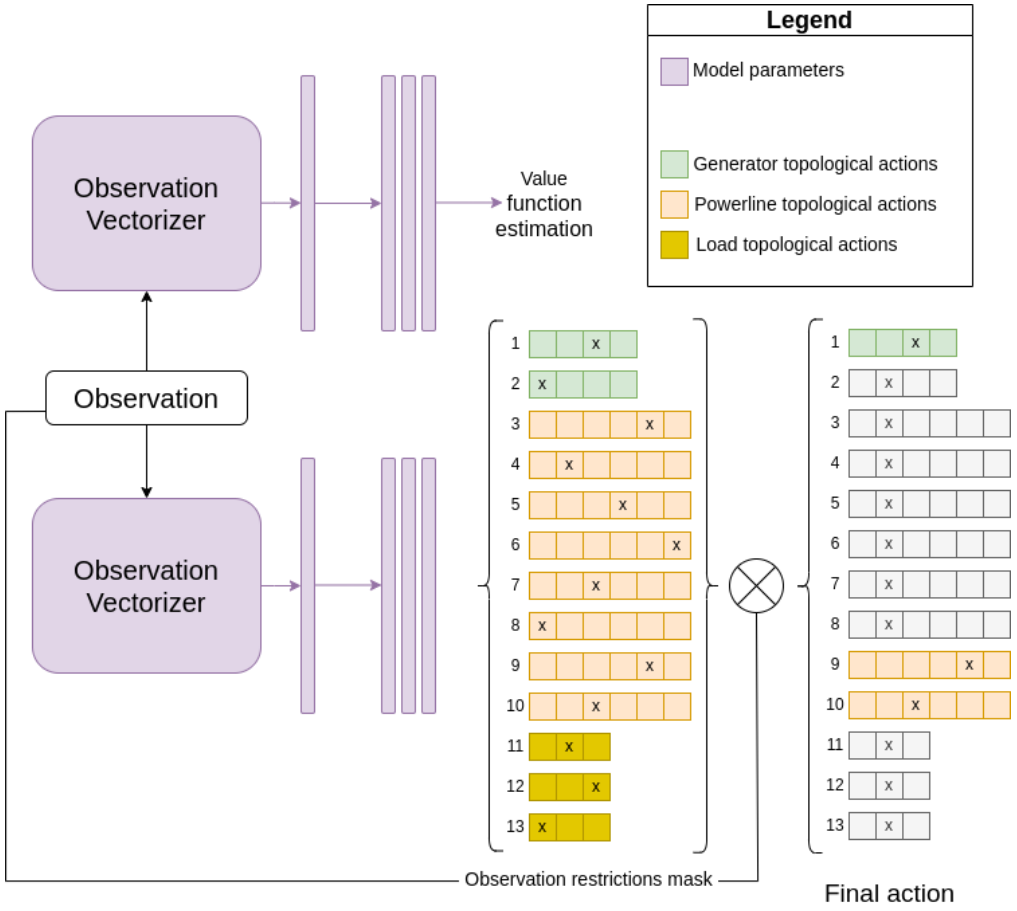
2. Metodología: PPO con máscaras

- Método Actor-Critic
  - El mismo agente dicta las acciones y evalúa lo "bueno" que es un estado para su objetivo y, por tanto, la acción más reciente.



- Variante con máscaras:
  - Nos permite establecer una lógica a través de la cual impedir acciones ante determinados escenarios.

2. Metodología: Espacio de acciones



- Una dimensión para cada elemento de la red:
  - 2 generadores (off, stay, bus 1, bus 2)
  - 3 cargas (stay, bus 1, bus 2)
  - 6 líneas (off, stay, + 4 combinaciones de barras)
- Permite fijar elementos a "stay"
- Imposibilita acciones ilegales.

2. Metodología: Recompensa

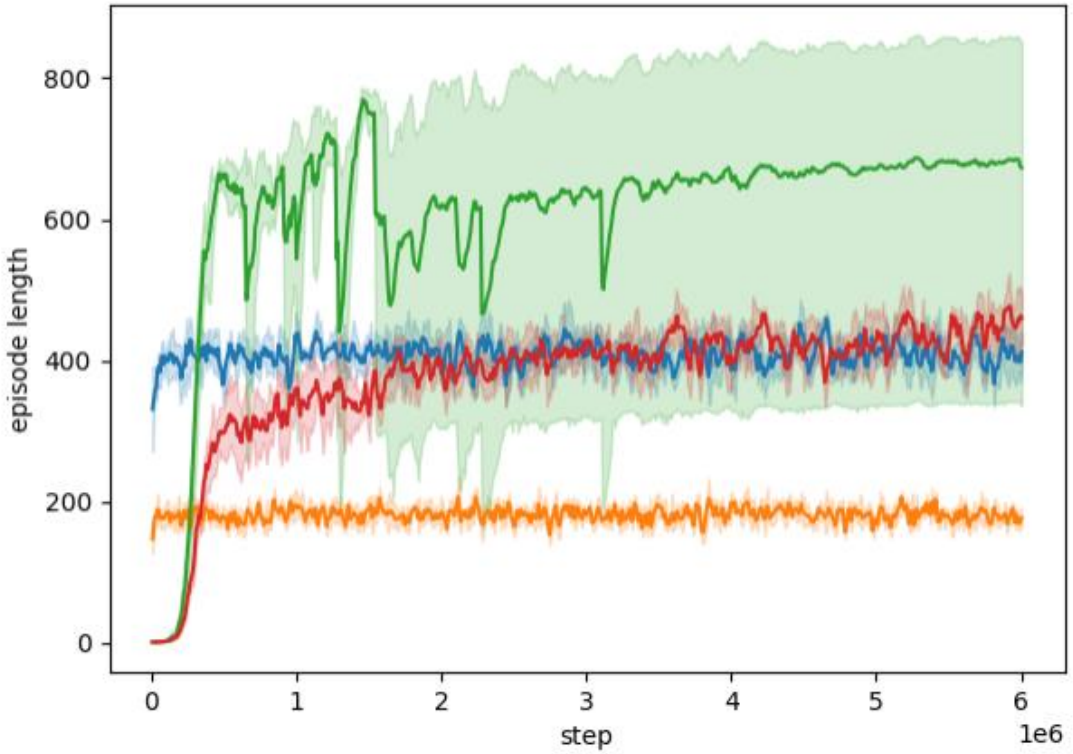
- Recompensa para un problema de control:
  - Positiva.
  - Favorece la longitud del episodio.
- Acotada.
- Reduce las pérdidas de energía.

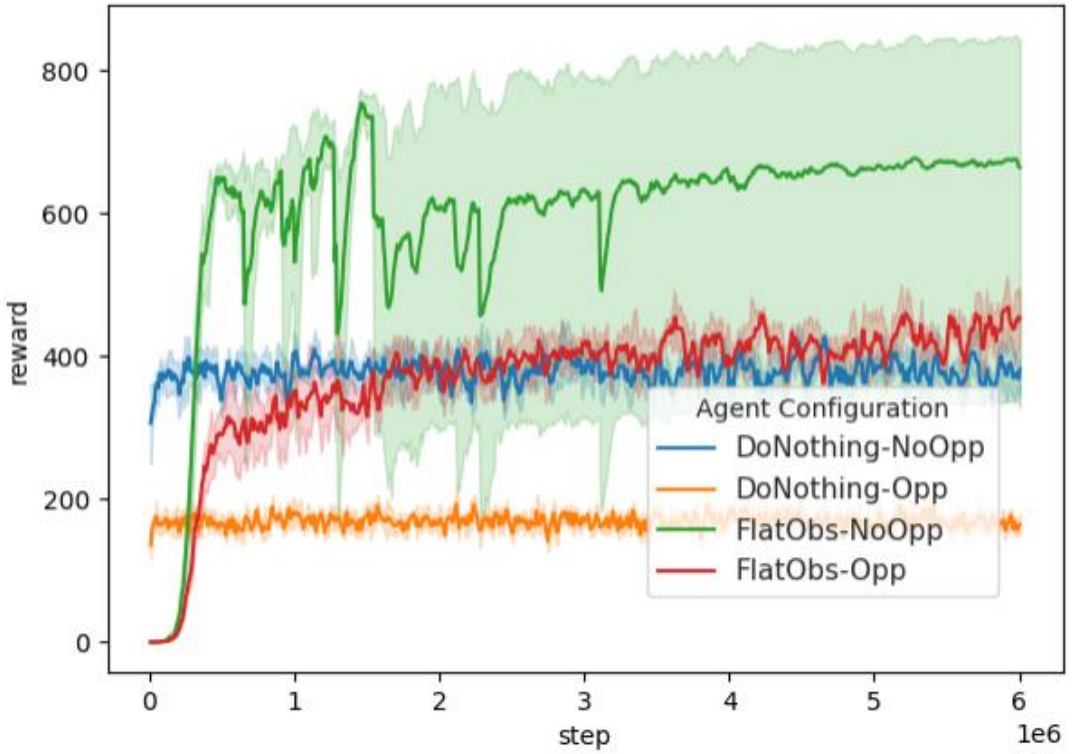
$$R(s_t, a_{t-1}) = \begin{cases} \frac{Loss_{max}(s_t) - Loss_{act}(s_t, a_{t-1})}{Loss_{max}(s_t)}, & \text{si } Loss_{max}(s_t) \neq 0 \\ 1, & \text{si } Loss_{max}(s_t) = 0 \end{cases}$$

2. Metodología: Configuración de entrenamiento

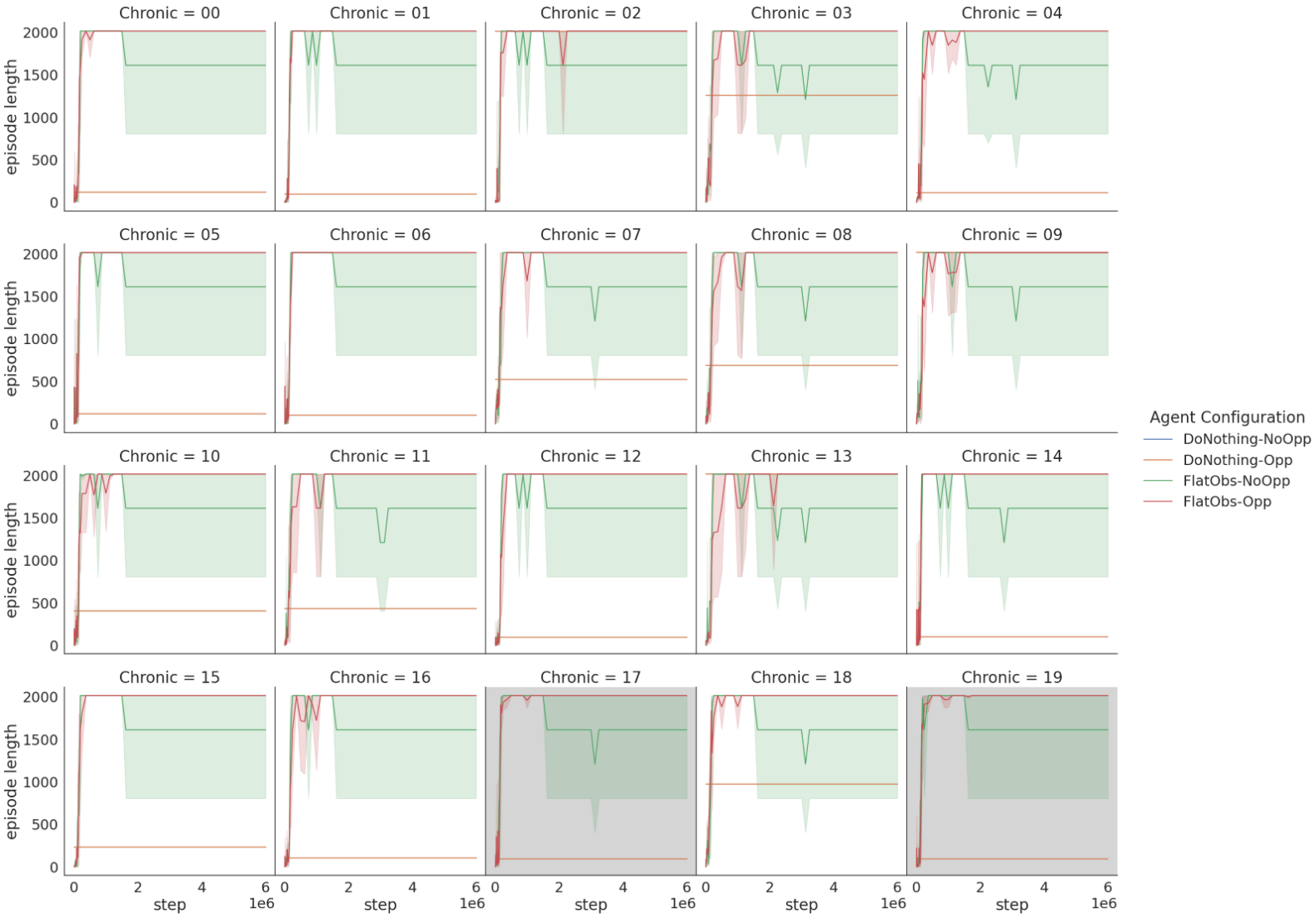
- Datos:
  - 18 crónicas x 5 puntos de arranque = 90 episodios
- Variantes:
  - Sin adversario.
  - Con adversario aleatorio
- Parámetros clave:
  - Paso temporal: 5 min
  - Truncado en 864 pasos (3 días)
  - 5 réplicas independientes para análisis de varianza.
  - Entrenamiento durante 6M de pasos.

3. Resultados: Entrenamiento





3. Resultados: Evaluación



4. Conclusiones y trabajo futuro

- Conclusiones
  - Es posible generar agentes robustos a través del aprendizaje por refuerzo
  - Entrenar con adversario genera agentes más robustos.

- Trabajo futuro:
  - Escalado a modelos IEEE-14 y IEEE-118 bus
  - Uso de grafos como representación de estado más rica.
  - Generación de modelos dinámicos para entrenamientos multi-red.

Preprint: **Graph-Enhanced Model-Free Reinforcement Learning Agents for Efficient Power Grid Topological Control**

III EDICIÓN CONGRESO DE
REDES INTELIGENTES
III WEBINAR 11 DE JUNIO 2025 - FUTURED

Gracias por su atención