Despacho Económico mediante Aprendizaje Reforzado: Un enfoque basado en Inteligencia Artificial

Diana Aidet Del Razo Ojeda David Romero Romero
Programa de Posgrado en Ingeniería Eléctrica, SEPI ESIME Zacatenco
Instituto Politécnico Nacional (IPN)
ddelrazoo1601@alumno.ipn.mx dromero@ipn.mx
https://orcid.org/0009-0003-2030-7480

***Resumen*— En la operación del sistema eléctrico de potencia el despacho económico es un problema clave, ya que su objetivo es minimizar los costos de energía, satisfacer la demanda y cumplir con las restricciones operativas de las unidades de generación. En este trabajo se propone el algoritmo del aprendizaje por refuerzo Q-learning para resolver el despacho de un sistema que está compuesto por plantas termoeléctricas y plantas de ciclo combinado con un solo nodo de demanda. Q-learning permite que un agente aprenda a asignar potencias por medio de estrategias optimas sin necesidad de tener un modelo explícito del sistema, a diferencia de los métodos tradicionales.**

Basado en el costo total y el cumplimento de la demanda se implementa un esquema de entrenamiento donde el agente ajusta las potencias de generación en función de una señal de recompensa. Se analiza los efectos de los parámetros del algoritmo tales como, tasa de aprendizaje, factor de descuento, la exploración-explotación. De acuerdo con los resultados que se muestran basado en el enfoque Q-learning encuentra configuraciones de generación que reducen el costo operativo y a su vez cumplen con las restricciones del sistema en este caso la demanda y los limites superiores e inferiores de los generadores, se demuestra una alternativa para la solución del despacho económico.

***Palabras Clave* — aprendizaje reforzado, despacho económico, optimización de sistemas eléctricos, Q-learning.**

I. Introducción

El despacho económico es un problema fundamental en la operación del sistema eléctrico de potencia, su objetivo es minimizar el costo de generación mientras satisface la demanda y cumple las restricciones técnicas de las unidades generadoras. [1]

Tradicionalmente, este problema se ha abordado por medio de técnicas de optimización clásica, como programación lineal, Newton Raphson y otras técnicas metaheurísticas, sin embargo, debido al gran avance del aprendizaje automático, ha surgido nuevas alternativas basada en la inteligencia artificial. En este contexto, el aprendizaje por refuerzo (Reinforcement Learning, RL) se ha comprobado que es una herramienta eficaz para la toma de decisiones en problemas dinámicos y complejos, donde desempeñan un papel crucial la incertidumbre y la exploración. [2] [3] A diferencias de los enfoques clásicos, RL permite que un agente aprenda estrategias optimas por medio de una iteración directa con el ambiente, así mejora su desempeño a lo largo del tiempo mediante recompensas y penalizaciones. [2]

Uno de los algoritmos más utilizados en este campo es el Q-learning, un método de aprendizaje libre de modelos que permite encontrar una política óptima de decisión sin necesidad de conocer exactamente la dinámica del sistema. [2] [4] [5] [6] En el problema del despacho económico, este enfoque promete una alternativa flexible para la asignación de generación entre múltiples plantas, donde aprenden estrategias de minimice el costo, pero al mismo tiempo sin comprometer el suministro de energía. [3]

En este trabajo se presenta la aplicación del Q-learning en el despacho económico de un sistema uninodal con múltiples plantas, donde se evalúa el desempeño en términos de costo, convergencia y estabilidad de la solución. Se analizan los parámetros fundamentales del algoritmo, su proceso de entrenamiento y los resultados obtenidos tras el aprendizaje del agente. Por último, se compara el rendimiento de Q-learning con enfoques tradicionales para determinar su factibilidad como un método de optimización en la operación de sistemas eléctricos de potencia.

II. Fundamentos Teóricos

Despacho Económico

El problema clásico del Despacho Económico se formula de la siguiente manera [7]:

Minimizar:

$$F = \sum_{i=1}^{n} f_i(P_i) \tag{1}$$

Sujeto a:

$$\sum_{i=1}^{n} P_i = P_D \rightarrow g(P_i) = \sum_{i=1}^{n} P_i - P_D = 0 \tag{2}$$

$$P_i \geq \underline{P_i} \qquad P_i \leq \overline{P}_i \tag{3}$$

Donde:

P_i : generación de la unidad i
f_i : costo de la unidad i
F : costo total de n unidades
P_D : demanda total
$\underline{P_i}$: límite inferior de la generación de la unidad i
$\overline{P_i}$: límite superior de la generación de la unidad i
n : número total de unidades

Aprendizaje reforzado

El aprendizaje reforzado es un enfoque dentro del aprendizaje automático donde un agente aprende y mejora su desempeño al interactuar con su ambiente por medio de un proceso de prueba y error. En este modelo, el agente toma decisiones en el estado actual del ambiente, recibe retroalimentación por medio de recompensa o penalizaciones. [2] [5] [8] [9]

Los agentes del aprendizaje reforzado se pueden clasificar en dos principales categorías [2]:

- **Basado en modelos**: Estos agentes obtienen o cuentan con un modelo de transición del entorno $P(s, a, s')$ y aprende una función de utilidad $U(s)$. Esta perspectiva permite planificar acciones a futuro y mejorar la toma de decisiones que se basa en la dinámica del entorno.
- **Sin modelo**: Para este caso el agente no aprende ni conoce un modelo de transición. Sino que aprende de una manera más eficiente de comportarse, esto adopta dos enfoques:
    1. Aprendizaje de utilidad de acciones.
    2. Búsqueda de políticas.

III. METODOLOGÍA

En este trabajo, se utiliza un agente sin modelo con aprendizaje de utilidad de acciones, llamado Q-learning para la solución del despacho económico de sistema con seis centrales eléctricas. El objetivo es minimizar el costo de generación con una demanda determinado, utilizando la técnica que emplea el aprendizaje reforzado.

Definición del problema

El sistema está conformado por seis unidades generadoras, cada una de ellas con una capacidad mínima y máxima de generación. En la **Tabla 1** se presentan las características técnicas, donde la planta 1,2,4 y 6 corresponden a plantas de ciclo combinado [10] [11] operando en su estado óptimo (estado 3), las cuales han sido modeladas por medio de un polinomio de cuarto grado. De forma complementaria, las unidades 3 y 5 corresponde a plantas termoeléctricas. El requerimiento total de demanda a cubrir, concentrada en un único nodo, es de 500 MW.

Tabla 1.- Parametros del sistema de prueba

Planta	Nodo	P_{min}	P_{max}	Coeficientes				
		(MW)	(MW)	x^4	x^3	x^2	x	c
1	1	95	295	0.0000094106	-0.0080	2.4259	-281.4737	15940
2	3	95	295	0.0000094106	-0.0080	2.4259	-281.4737	15940
3	4	10	60	0	0	0.00895	18.3538	181.2980
4	5	30	135	-0.0000074968	0.0023	-0.2135	18.9162	388.3228
5	6	10	60	0	0	0.00664	30.4000	197.0575
6	7	30	135	-0.0000074968	0.0023	-0.2135	18.9162	388.3228

Parámetros del algoritmo de Q-learning

Para el entrenamiento del agente, se han establecido los siguientes parámetros:

Factor de aprendizaje (α) : 0.05, se ajusta dinámicamente durante la fase de explotación.
Factor de descuento (γ) : 0.995, se utiliza para valorar la relevancia de las recompensas futuras.
Tasa de exploración inicial (ε) : 1.0, disminuye progresivamente hasta alcanzar un valor mínimo de 0.01.
Episodios totales considerados en el entrenamiento: 10000000.

Estructura del aprendizaje

El algoritmo de Q-learning opera mediante la actualización de una tabla de valores $Q(s, a)$ de forma iterativa, donde [2]:

s : representa el estado de sistema, para este caso determina la generación total de las plantas.
a : corresponde a la acción seleccionada, en este caso al ajuste de la potencia de una unidad generadora.
La tabla Q registra la recompensa esperada vinculada a cada combinación de estado y acción

Fase de entrenamiento

El proceso de entrenamiento se desarrolla siguiendo los pasos que se describen a continuación:

1. Se establece un estado inicial de manera aleatoria dentro de los límites de generación de cada planta.
2. La potencia de salida de cada planta se ajusta para asegurar el cumplimiento de la demanda previo a la evaluación de costo.
3. Se evalúa el costo de generación total a partir de la función objetivo, la cual representa los costos operativos de cada planta.
4. Se establece la recompensa en función del desbalance entre la potencia generada y la demanda requerida:
    - Si la diferencia es inferior a 1 MW, se otorga un valor positivo, la cual depende de la minimización del costo de operación.
    - Si la diferencia es mayor, se penaliza al agente con un término que es proporcional al error cubico de la diferencia.
5. Se actualiza la tabla Q con la ecuación [6]:

$$Q(s,a) = (1-\alpha)Q(s,a) + \alpha\left[r + \gamma \max_{a'} Q(s',a')\right] \quad (5)$$

6. El valor de α se ajusta de manera dinámica con el objetivo de mejorar la estabilidad al concluir el entrenamiento.
7. Se emplea una estrategia de exploración-explotación para incrementarla efectividad del aprendizaje.
   - Con la probabilidad ε, el agente elige una acción aleatoria (exploración).
   - De lo contrario, selecciona la mejor acción determinada en función de la tabla Q.
8. Gradualmente se reduce ε con el fin de favorecer la explotación de las últimas iteraciones.

Fase de evaluación y resultados

Una vez finalizado el entrenamiento, se identifica la solución óptima, caracteriza por cumplir con los siguientes criterios:

- Cumple con la demanda total con un margen de error inferior a 1 MW.
- Minimiza el costo total de generación.

Se lleva a cabo una visualización de la tabla Q que se obtiene mediante un mapa de calor, lo que permite analizar la convergencia del algoritmo y la estrategia de decisión del agente.

IV. Resultados y Discusión

Evaluación del rendimiento del algoritmo

Tras el entrenamiento de 10000000 episodios, el agente Q-learning consiguió determinar configuraciones óptimas de generación que minimizan el costo total cumpliendo con la demanda de 500 MW. La mejor solución alcanzada presenta una distribución de potencia entre las plantas que respeta sus límites de operación.

Se analizan los siguientes aspectos para evaluar el rendimiento del algoritmo:

- Mapa de calor de la tabla Q.
- Distribución de potencias.
- Comparación con otros métodos

Mapa de calor de la tabla Q

Con el propósito de examinar como aprendió el agente a asociar los estados con las mejores acciones, se presenta un mapa de calor de la tabla Q en la Figura 2. En el gráfico, el eje vertical corresponde a los **estados**, se definen en función de distintos niveles de generación, el eje horizontal representa las **acciones**, asociadas a la operación de diferentes plantas generadoras.

Se identifica que los valores de la tabla Q de manera comparativa son más bajos en la mayoría de los estados, lo cual indica que las acciones no contribuyen de manera significativa a la mejora de la recompensa acumulada durante el proceso de entrenamiento, por otro lado, con excepción notable al estado 10 demuestra que, para dicho estado, las acciones correspondientes registran valores significativamente en la tabla Q , esto evidencia que dichas decisiones han sido consideradas las más adecuadas para alcanzar la máxima recompensa posible esperada en esa etapa del aprendizaje.

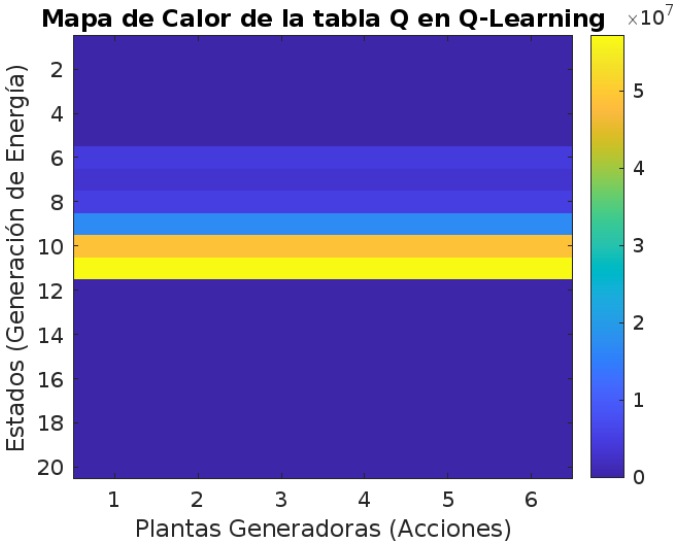

Figura 2.- Mapa de calor de la tabla Q después del entrenamiento.

Distribución óptima de potencias

En la Figura 3, se presenta la distribución final de generación correspondiente a las seis plantas. Se observa que la asignación de generación respeta los límites de capacidad de cada unidad, satisfaciendo la totalidad de la demanda del sistema.

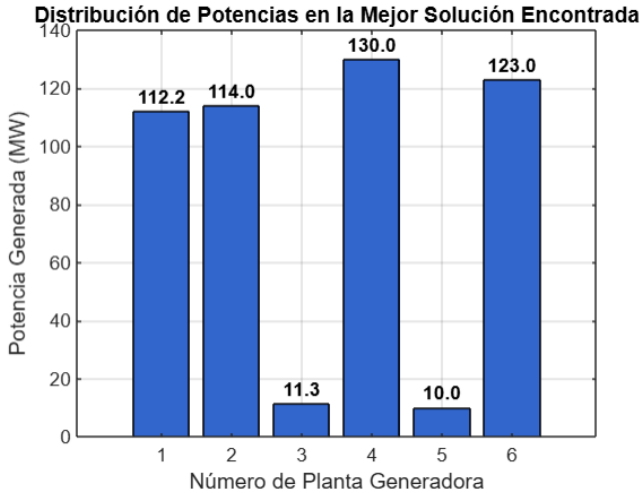

Figura 3.- Distribución de potencia de las plantas generadoras en la mejor solución encontrada.

Comparación con otros métodos

Con el propósito de evaluar la efectividad de Q-learning, se presenta una comparación de los costos totales en distintos métodos en la **Tabla 2**.

Tabla 2.- Comparación de costos

Método	Costo total ($)	CPU time (s)
Q-learning	15 296	30.8770
Evolución Diferencial (ED)	15 339	0.6821
ED mejorada	15 277	0.7518
Programación No Lineal (PNL)	15 277	1.0308

De acuerdo con los resultados presentados en la **Tabla 2**, se puede observar que la ED mejorada y la PNL obtuvieron los menores costos con $15277, esto demuestra que ambos métodos son altamente efectivos y consistentes para resolver este tipo de problemas.

Por otro lado, el método de Q-learning, mostro un costo competitivo de $15296, se ubicó por encima de los costos de los métodos anteriormente mencionados, evidenciando la necesidad de ajustes adicionales en sus parámetros de aprendizaje (tasa de exploración, tasa de aprendizaje y factor de descuento).

Sin embargo, es importante destacar que Q-learning tiene ventajas estratégicas, como su capacidad de apartarse en tiempo real a cambios en el sistema y operar sin un modelo matemático preciso del problema, lo que resulta ventajoso en entornos dinámicos o inciertos.

V. Conclusiones

En este trabajo se simulo y evaluó el algoritmo de Q-learning para la solución del despacho económico en un sistema eléctrico. Por medio del aprendizaje reforzado, el agente fue capaz de ajustar la asignación de potencia de las unidades generadoras con el objetivo de minimizar el costo de operación y cumplir con la demanda determinada.

Los resultados que se obtienen demuestran que Q-learning es una alternativa viable para este tipo de problemas, se alcanzó convergencia hacia soluciones eficientes sin necesidad de un modelo explícito del sistema. Se observo que el desempeño del algoritmo depende directamente de la configuración de sus parámetros como tasa de aprendizaje, el factor de descuento y la estrategia de exploración-explotación. Una correcta parametrización contribuyó a incrementar la estabilidad del entrenamiento y optimizar la calidad de la solución obtenida.

Si bien el algoritmo demostró resultados prometedores, hay aspectos que podrían ser mejorados en trabajos posteriores. La implementación de técnicas avanzadas como Deep Q-Networks (DQN) o el uso de modelos híbridos podrían facilitar una optimización más sólida robusta en sistemas de gran escala. De la misma manera, la incorporación de restricciones adicionales, como, apagado de unidades generadoras y costos de arranque, permitirá una mayor aproximación del modelo en condiciones operativas reales de los sistemas eléctricos.

Agradecimientos

Los autores agradecen al Instituto Politécnico Nacional (IPN) por el apoyo brindado por medio de la licencia de MATLAB utilizada en este trabajo, así como, a la Secretaría de Ciencia, Humanidades, Tecnología e Innovación (SECIHTI) por la beca otorgada para la realización de este estudio.

Referencias

[1] A. J. Wood, B. F. Wollenberg y G. B. Sheblé, Power generation, operation and control, Wiley, 2013.

[2] S. Russel y P. Norving, Artificial Intelligence A Mordern Approach, Pearson , 2022.

[3] E. Jasmin, T. Ahamed y V. Jagathiraj, «A Reinforcement Learning Algorithm to Economic Dispatch Considering Transmission Losses,» *TENCON 2008 - 2008 IEEE Region 10 Conference*, pp. 1-6, 2008.

[4] H. T., H. Tang y S. Levine, «Reinforcement learning with deep energy-based polices,» *Proc. of the 34th international conference on machine learning*, pp. 1352-1361, 2017.

[5] Z. Zidong, Z. Dongxia y R. C. Qui, «Deep Reinforcement Learning for Power System Applications: An Overview,» *CSEE Journal of power and energy systems* , vol. 6, nº 1, 2020.

[6] C. J. Watkins y P. Dayan, «Q-learning,» *Machine Learning*, vol. 8, nº 3-4, pp. 279-292, 1992.

[7] F. Gao y S. G. B., «Stochastic Optimization Techniques for Economic Dispatch with Combined Cycle Units,» *9th International Conference on Probabilistic Methods Applied to Power Systems*, 2006.

[8] L. Pawel, W. Lilian, K. Minwoo y O. Hyondong, «Exploration in deep reinforcement learning: A survey,» *Information Fusion*, vol. 85, pp. 1-22, 2022.

[9] R. S. Sutton y A. G. Barto, Reinforcement Learning: An Introduction, 2015.

[10] L. Bayón, G. García Nieto, M. Ruiz y P. Suarez, «An economic dispatch algorithm of combined cycle units,» *International Journal of Computer Mathematics*, vol. 91, nº 2, pp. 269-277, 2014.

[11] J. Alemany, D. Moitre, H. Pinto y M. F., «Short-Term Scheduling of Combined Cycle Units Using Mixed Integer Linear Programming Solution,» *Energy and Power Engineering*, pp. 161-170, 2013.