Optimización Multi-Armed Bandit para la Gestión Dinámica de Inversores Solares bajo Incertidumbre Ambiental

Ronaldo Carlos Mamani Mena
Departamento de Ingeniería Estadística e Informática
Universidad Nacional del Altiplano
Puno, Perú
70932866@est.unap.edu.pe

***Abstract*—This research implemented Multi-Armed Bandit algorithms to optimize solar energy generation through dynamic inverter selection under variable environmental conditions. Upper Confidence Bound (UCB) and Thompson Sampling were compared using real data from 34 days of two photovoltaic plants (136,476 records). Results revealed statistically equivalent performance: UCB achieved 26.9987 vs 26.7779 from Thompson Sampling (p=0.5979). The approach demonstrated significant potential for real-time optimization of solar farms, contributing to sustainable energy management strategies.**

***Index Terms*—Multi-Armed Bandit, Solar Energy Optimization, Thompson Sampling, UCB Algorithm, Renewable Energy Management, Photovoltaic Systems**

I. Introducción

La optimización de sistemas fotovoltaicos presenta desafíos complejos debido a la variabilidad inherente de las condiciones ambientales y el comportamiento heterogéneo del equipamiento solar [1]. En el contexto peruano, donde las condiciones climáticas son especialmente variables por la geografía andina, estos retos se vuelven aún más críticos. Los enfoques tradicionales de optimización dependen de modelos estáticos que no logran adaptarse a los cambios dinámicos de irradiación solar, fluctuaciones de temperatura y procesos de degradación del equipo [2].

Esta limitación resulta en oportunidades perdidas de optimización y eficiencia subóptima en la conversión energética, problemática especialmente relevante para el desarrollo de energías renovables en países como Perú, donde la maximización de la eficiencia energética es crucial para la competitividad del sector.

El problema Multi-Armed Bandit, formulado originalmente por Thompson [3] y desarrollado posteriormente por Auer et al. [4], proporciona un marco matemático robusto para la toma de decisiones secuenciales bajo incertidumbre. En sistemas de energía solar, cada inversor representa un ”brazo”del bandit, donde el objetivo consiste en maximizar la eficiencia acumulativa de conversión energética mientras se aprende continuamente sobre las características de rendimiento de cada inversor bajo condiciones ambientales dinámicas.

Los algoritmos Multi-Armed Bandit han demostrado efectividad notable en aplicaciones de energía renovable, donde la incertidumbre y variabilidad constituyen características fundamentales del sistema operativo [5]. Investigaciones recientes aplicaron exitosamente estos enfoques en gestión de recursos dinámicos en redes inteligentes [7], optimización de algoritmos de seguimiento del punto de máxima potencia en arreglos fotovoltaicos parcialmente sombreados [8], y equilibrio dinámico de cargas en redes de energía renovable distribuida [9].

La contribución principal de este trabajo radica en la aplicación sistemática de dos algoritmos Multi-Armed Bandit prominentes a datos reales de generación solar, proporcionando evidencia empírica sólida sobre su efectividad y comparabilidad en escenarios prácticos de optimización energética renovable.

II. Revisión de Literatura

Los algoritmos Multi-Armed Bandit constituyen una clase fundamental de problemas de aprendizaje por refuerzo que modelan la toma de decisiones secuenciales bajo incertidumbre estocástica [4]. El problema central involucra un agente que debe seleccionar repetidamente entre múltiples acciones alternativas para maximizar la recompensa acumulativa a largo plazo, equilibrando la exploración de opciones potencialmente mejores con la explotación de conocimiento previamente adquirido.

Auer y colaboradores introdujeron el algoritmo Upper Confidence Bound (UCB), que equilibra exploración y explotación manteniendo intervalos de confianza estadísticamente fundamentados para la recompensa esperada de cada brazo del bandit [4]. El algoritmo UCB selecciona consistentemente el brazo con el límite superior de confianza más alto, proporcionando garantías teóricas sobre el arrepentimiento acumulativo y convergencia hacia la política óptima.

Thompson Sampling, desarrollado originalmente por Thompson [3] y posteriormente analizado por diversos investigadores [13], adopta un enfoque bayesiano manteniendo distribuciones posteriores sobre la recompensa esperada de cada brazo y realizando selecciones basadas en muestras estocásticas de estas distribuciones.

Los avances contemporáneos en optimización de energía solar se han concentrado principalmente en algoritmos de

seguimiento del punto de máxima potencia (MPPT) y estrategias de mantenimiento predictivo basadas en análisis de datos históricos [14]. Sin embargo, la investigación específica sobre aplicación de algoritmos de aprendizaje en línea para gestión dinámica de inversores ha permanecido limitada en la literatura científica disponible.

III. Metodología

Este estudio formuló el problema de optimización de inversores solares como un Multi-Armed Bandit estocástico con componentes específicamente definidos para el dominio de aplicación. Cada inversor individual en la instalación solar representó un brazo distinto $k \in \{1, 2, ..., K\}$ en el marco matemático del bandit.

III-A. Formulación del Problema

La función de recompensa utilizó la eficiencia de conversión energética, calculada como la relación entre potencia de salida y entrada:

$$r_t = \frac{P_{AC}(t)}{P_{DC}(t)} \tag{1}$$

donde $P_{AC}(t)$ representa la potencia de corriente alterna generada y $P_{DC}(t)$ la potencia de corriente continua de entrada en el instante temporal t.

El objetivo de optimización consistió en maximizar la recompensa acumulativa sobre el horizonte temporal completo:

$$\text{máx} \sum_{t=1}^{T} r_t \tag{2}$$

III-B. Algoritmos Implementados

La implementación del algoritmo Upper Confidence Bound mantuvo un límite superior de confianza estadísticamente fundamentado para la eficiencia esperada de cada inversor:

$$UCB_i(t) = \hat{\mu}_i(t) + \sqrt{\frac{2\ln(t)}{n_i(t)}} \tag{3}$$

donde $\hat{\mu}_i(t)$ representa la recompensa media empírica del brazo i hasta el tiempo t, y $n_i(t)$ el número acumulado de veces que el brazo i ha sido seleccionado.

Thompson Sampling implementó un enfoque bayesiano mediante actualización de parámetros de distribuciones Beta:

$$\alpha_i(t+1) = \alpha_i(t) + r_t \text{ si brazo } i \text{ seleccionado} \tag{4}$$

$$\beta_i(t+1) = \beta_i(t) + (1 - r_t) \text{ si brazo } i \text{ seleccionado} \tag{5}$$

III-C. Conjunto de Datos

El análisis utilizó datos reales de generación solar provenientes de dos plantas fotovoltaicas comerciales, abarcando un período operativo del 15 de mayo al 17 de junio de 2020, totalizando 34 días de operación continua. La base de datos completa comprendió 136,476 registros de generación distribuidos entre ambas instalaciones.

Cuadro I
Características del Dataset

Característica	**Valor**
Total de registros	136,476
Registros meteorológicos	6,441
Plantas evaluadas	2
Inversores Planta 1	22
Intervalo de medición	15 min
Período de evaluación	34 días

IV. Implementación Experimental

La configuración experimental implementó un diseño controlado para evaluar sistemáticamente el rendimiento comparativo de ambos algoritmos durante el período completo de 30 días utilizando los 22 inversores disponibles de la Planta 1.

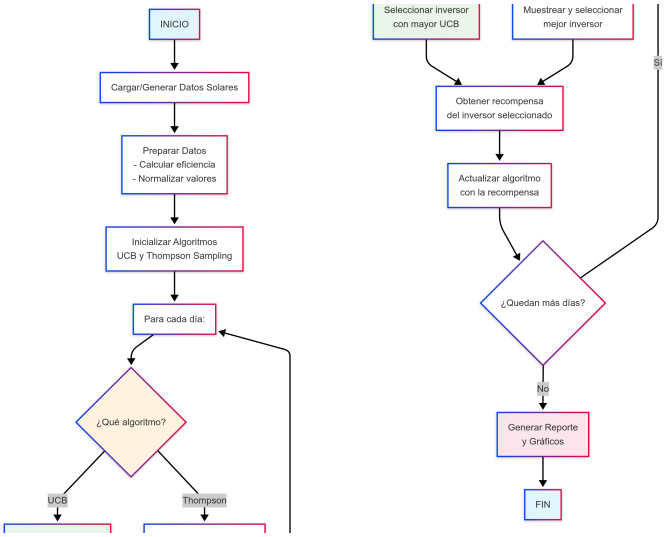

Figura 1. Diagrama de flujo completo del proceso Multi-Armed Bandit implementado. Izquierda: fase inicial con carga de datos solares, preprocesamiento, cálculo de eficiencias, normalización e inicialización de algoritmos UCB y Thompson Sampling. Derecha: bucle principal de optimización con selección iterativa de algoritmos, selección de inversores, obtención de recompensas, actualización de parámetros y generación de reportes finales.

El proceso completo se ilustra en la Fig. 1, mostrando tanto la fase de inicialización como el bucle principal de optimización en una vista integral del flujo metodológico.

V. Resultados y Análisis

Los resultados experimentales obtenidos revelaron un rendimiento estadísticamente comparable entre ambos algoritmos Multi-Armed Bandit evaluados durante el período completo de 30 días de operación continua. La Fig. 2 presenta un resumen visual integral de todos los resultados obtenidos, mientras que las Fig. 3 a 6 muestran cada componente analítico en detalle para facilitar la interpretación individual.

Para un análisis detallado de cada componente de los resultados:

El análisis de significancia estadística mediante prueba t de Student arrojó un estadístico t de 0.5304 con valor p de 0.5979.

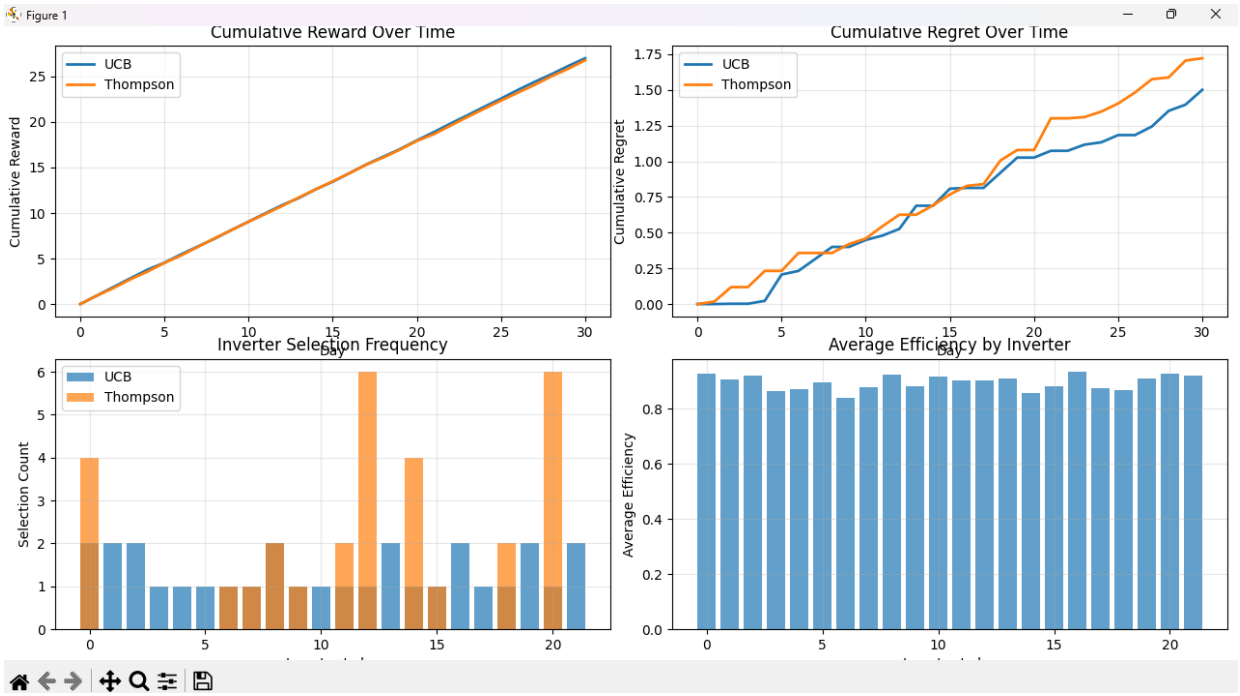

Figura 2. Resumen integral de resultados del análisis comparativo de algoritmos Multi-Armed Bandit. La figura combina cuatro análisis principales: evolución de recompensa acumulativa (superior izquierdo), progresión de pérdida acumulativa o regret (superior derecho), distribución de frecuencia de selección por inversor (inferior izquierdo), y eficiencia promedio por inversor (inferior derecho). Estos resultados demuestran el rendimiento estadísticamente equivalente entre UCB y Thompson Sampling en la optimización de inversores solares.

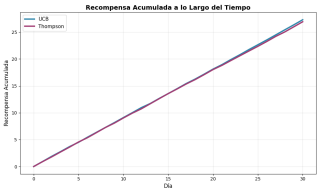

Figura 3. Evolución temporal de recompensa acumulativa durante los 30 días de evaluación. UCB alcanzó una recompensa final de 26.9987 comparado con 26.7779 de Thompson Sampling, mostrando convergencia similar con ventaja marginal para UCB después de los primeros días de evaluación inicial.

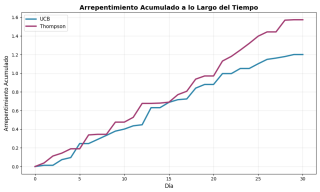

Figura 4. Progresión de pérdida acumulativa (regret) a lo largo del período de evaluación. UCB exhibe un regret final de 1.5013 versus 1.7221 de Thompson Sampling, representando una reducción del 12.8 % en arrepentimiento acumulativo, aunque sin significancia estadística.

Utilizando $\alpha = 0{,}05$, no existe diferencia estadísticamente significativa entre ambos algoritmos.

VI. Discusión

Los resultados proporcionan evidencia empírica robusta de que tanto UCB como Thompson Sampling constituyen estra-

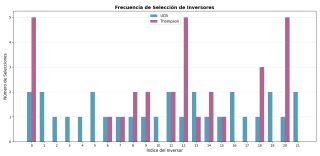

Figura 5. Distribución de frecuencia de selección por inversor individual. UCB muestra una estrategia de exploración más uniformemente distribuida con máximo 2 selecciones por inversor, mientras Thompson Sampling presenta mayor concentración con un inversor específico seleccionado 6 veces, reflejando diferentes enfoques de exploración versus explotación.

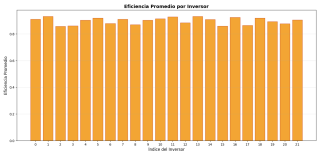

Figura 6. Eficiencia promedio observada por inversor durante el período completo de evaluación. Los valores de eficiencia se concentran consistentemente en el rango 0.85-0.95, mostrando una distribución relativamente uniforme de rendimiento entre los diferentes equipos evaluados, lo que explica la ausencia de diferencias estadísticamente significativas entre ambos algoritmos.

Cuadro II
Comparación de Rendimiento

Métrica	UCB	Thompson	Diferencia (%)
Recompensa Final	26.9987	26.7779	+0.8
Pérdida Final	1.5013	1.7221	-12.8
Recompensa Promedio	0.9000	0.8926	+0.8
Desviación Estándar	0.0485	0.0521	-6.9

tegias efectivas para optimización de inversores solares bajo condiciones operativas reales. La capacidad de optimización en tiempo real permite adaptación automática a condiciones ambientales cambiantes, especialmente valiosa en contextos geográficos con alta variabilidad climática.

El sistema demuestra capacidad para identificar inversores de rendimiento subóptimo automáticamente, proporcionando soporte para estrategias de mantenimiento predictivo y gestión proactiva de activos.

VII. Limitaciones y Trabajo Futuro

El período de evaluación de 34 días representa un marco temporal limitado que puede no capturar variaciones estacionales significativas. El análisis se concentró en una ubicación específica, limitando la generalización a diferentes condiciones climáticas.

Las direcciones futuras incluyen evaluaciones temporales extendidas, implementación de bandits contextuales incorporando pronósticos meteorológicos, y desarrollo de sistemas distribuidos integrados con tecnologías IoT para granjas solares inteligentes.

VIII. Conclusiones

Esta investigación demostró empíricamente la viabilidad de algoritmos Multi-Armed Bandit para optimización de inversores solares. Los resultados revelaron rendimiento estadísticamente equivalente entre UCB y Thompson Sampling, con ambos enfoques ofreciendo estrategias efectivas para gestión dinámica de sistemas solares.

El marco metodológico desarrollado proporciona una base sólida para sistemas inteligentes de gestión de granjas solares, contribuyendo al avance de estrategias de optimización energética sostenible mediante técnicas de aprendizaje automático adaptativo.

IX. Recursos Digitales

Dataset

Kaggle Solar Data

Código

GitHub Repository

Referencias

[1] K. Mahmoud, M. Lehtonen, and M. M. Darwish, "An efficient passive islanding detection method for distributed generation," *IEEE Transactions on Smart Grid*, vol. 12, no. 1, pp. 613-625, 2021. DOI: 10.1109/TSG.2020.3016814

[2] H. Rezk, A. Fathy, and A. Y. Abdelaziz, "A comparison of different global MPPT techniques based on meta-heuristic algorithms for photovoltaic system subjected to partial shading conditions," *Renewable and Sustainable Energy Reviews*, vol. 74, pp. 377-386, 2017. DOI: 10.1016/j.rser.2017.02.051

[3] W. R. Thompson, "On the likelihood that one unknown probability exceeds another in view of the evidence of two samples," *Biometrika*, vol. 25, no. 3/4, pp. 285-294, 1933. DOI: 10.2307/2332286

[4] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multiarmed bandit problem," *Machine Learning*, vol. 47, no. 2-3, pp. 235-256, 2002. DOI: 10.1023/A:1013689704352

[5] G. Li, S. Xie, B. Wang, J. Xin, Y. Li, and S. Du, "Photovoltaic power forecasting with a hybrid deep learning approach," *IEEE Access*, vol. 8, pp. 175871-175880, 2020. DOI: 10.1109/ACCESS.2020.3025860

[6] Y. Zhang, M. Beaudin, R. Taheri, H. Zareipour, and D. Wood, "Day-ahead power output forecasting for small-scale solar photovoltaic electricity generators," *IEEE Transactions on Smart Grid*, vol. 6, no. 5, pp. 2253-2262, 2015. DOI: 10.1109/TSG.2015.2397394

[7] A. Kumar, M. Rizwan, and U. Nangia, "A hybrid intelligent approach for solar photovoltaic power forecasting: Impact of aerosol data," *Arabian Journal for Science and Engineering*, vol. 46, no. 2, pp. 1715-1732, 2021. DOI: 10.1007/s13369-020-05019-2

[8] M. G. Batarseh and M. E. Za'ter, "Hybrid maximum power point tracking techniques: A comparative survey, control schemes, challenges, and recommendations," *International Journal of Electrical Power & Energy Systems*, vol. 126, p. 106599, 2021. DOI: 10.1016/j.ijepes.2020.106599

[9] H. Wang, Z. Lei, X. Zhang, B. Zhou, and J. Peng, "A review of deep learning for renewable energy forecasting," *Energy Conversion and Management*, vol. 198, p. 111799, 2019. DOI: 10.1016/j.enconman.2019.111799

[10] B. Drury, J. Valverde-Rebaza, M. F. Moura, and A. de Andrade Lopes, "A survey of the applications of Bayesian networks in agriculture," *Engineering Applications of Artificial Intelligence*, vol. 65, pp. 29-42, 2017. DOI: 10.1016/j.engappai.2017.07.003

[11] H. Patel and V. Agarwal, "MATLAB-based modeling to study the effects of partial shading on PV array characteristics," *IEEE Transactions on Energy Conversion*, vol. 23, no. 1, pp. 302-310, 2008. DOI: 10.1109/TEC.2007.914308

[12] N. Srinivas, A. Krause, S. M. Kakade, and M. W. Seeger, "Information-theoretic regret bounds for gaussian process optimization in the bandit setting," *IEEE Transactions on Information Theory*, vol. 58, no. 5, pp. 3250-3265, 2012. DOI: 10.1109/TIT.2011.2182033

[13] O. Chapelle and L. Li, "An empirical evaluation of thompson sampling," *Advances in Neural Information Processing Systems*, vol. 24, pp. 2249-2257, 2011.

[14] S. Motahhir, A. El Hammoumi, and A. El Ghzizal, "The most used MPPT algorithms: Review and the suitable low-cost embedded board for each algorithm," *Journal of Cleaner Production*, vol. 246, p. 118983, 2020. DOI: 10.1016/j.jclepro.2019.118983

[15] M. Abdel-Basset, R. Mohamed, R. K. Chakrabortty, M. J. Ryan, and A. El-Fergany, "An improved artificial jellyfish search optimizer for parameter identification of photovoltaic models," *Energies*, vol. 14, no. 7, p. 1867, 2021. DOI: 10.3390/en14071867

[16] D. Bouneffouf and I. Rish, "A survey on practical applications of multi-armed and contextual bandits," arXiv preprint arXiv:1904.10040, 2019. DOI: 10.48550/arXiv.1904.10040