

Optimización Bayesiana de sistemas de recomendación en películas para minimizar el RMSE en predicción de ratings

Kenny Leonel Ccora Quispe

Facultad de Ingeniería Estadística e Informática, Universidad Nacional del Altiplano

Resumen

Los sistemas de recomendación modernos se enfrentan al reto de optimizar su rendimiento en conjuntos de datos masivos, garantizando al mismo tiempo una configuración eficiente del modelo. Este estudio examina la eficacia de la Optimización Bayesiana (BO) en la calibración de hiperparámetros del modelo SVD, en comparación con estrategias tradicionales como la Búsqueda en Cuadrícula y la Búsqueda Aleatoria, aplicadas al conjunto de datos MovieLens. Se realiza un análisis comparativo con un presupuesto de evaluación limitado, utilizando el RMSE como métrica de rendimiento. Los resultados demuestran que la BO supera consistentemente la minimización absoluta del RMSE, aunque no siempre alcanza una convergencia acelerada. Se presentan análisis detallados sobre la evolución de los métodos, las ventajas y desventajas entre la exploración y la explotación, y la eficiencia computacional de cada enfoque.

Palabras clave: Optimización Bayesiana, Sistemas de Recomendación, Evaluaciones Costosas, MovieLens, SVD.

Bayesian optimization of movie recommender systems to minimize RMSE in rating prediction

Abstract

Modern recommender systems face the challenge of optimizing their performance on massive datasets while ensuring efficient model configuration. This study examines the effectiveness of Bayesian Optimization (BO) in calibrating SVD model hyperparameters, compared to traditional strategies such as Grid Search and Random Search, applied to the MovieLens dataset. A comparative analysis is performed under a restricted evaluation budget, using RMSE as the performance metric. The results demonstrate that BO consistently outperforms absolute RMSE minimization, although it does not always achieve accelerated convergence. Detailed analyses are presented on the evolution of the methods, the trade-offs between exploration and exploitation, and the computational efficiency of each approach.

Keywords: Bayesian Optimization, Recommender Systems, Expensive Evaluations, MovieLens, SVD.

1. Introducción

El advenimiento de plataformas digitales que manejan grandes volúmenes de información ha impulsado el desarrollo de sistemas de recomendación capaces de ofrecer sugerencias personalizadas. En particular, los algoritmos basados en factorización matricial, como la descomposición en valores singulares (SVD), han demostrado ser altamente efectivos para capturar relaciones latentes entre usuarios e ítems, lo cual se refleja en su uso extensivo en el dataset MovieLens 20M (Harper y Konstan, 2015)[Harper and Konstan, 2015], donde se presentan millones de interacciones de calificación.

Sin embargo, para que estos modelos alcancen un rendimiento óptimo, es crítico optimizar los hiperparámetros, tales como el número de factores latentes y el grado de regularización. Cada combinación de hiperparámetros implica

entrenar el modelo por completo, lo que se traduce en un elevado costo computacional. Técnicas tradicionales como Grid Search y Random Search funcionan bien en espacios poco extensos, pero su rendimiento disminuye drásticamente cuando el número de combinaciones aumenta o cuando el tiempo para cada evaluación es significativo (Bergstra y Bengio, 2012)[Bergstra and Bengio, 2012].

Frente a esto, la Optimización Bayesiana (BO) emerge como una solución eficaz para optimizar funciones activas costosas. Este método construye una función aproximada probabilística (por ejemplo, un proceso gaussiano) y utiliza funciones de adquisición, como Expected Improvement, para seleccionar adaptativamente los próximos puntos de evaluación (Snoek et al., 2012)[Snoek et al., 2012]. La aplicación de BO en sistemas de recomendación ha demostrado su superioridad respecto a los métodos tradicionales, mostrando

mejoras en la elección de hiperparámetros como número de factores, tasa de aprendizaje y regularización, especialmente bajo un número limitado de evaluaciones (Galuzzi et al., 2020)[Galuzzi et al., 2020].

Este artículo propone un estudio comparativo entre Optimización Bayesiana, Random Search y Grid Search, limitando cada técnica a 40 evaluaciones. Se evalúa el impacto de cada estrategia en el desempeño del modelo SVD, utilizando como métricas el error cuadrático medio (RMSE) y el tiempo requerido para entrenar y evaluar el modelo. El objetivo es demostrar la eficiencia relativa de cada método en contextos donde cada evaluación es costosa, situación cada vez más común en aplicaciones de recomendación a gran escala.

2. Metodología

Para el desarrollo del modelo se utilizó el conjunto de datos MovieLens 20M, el cual contiene más de 20 millones de calificaciones explícitas realizadas por usuarios a películas. Este conjunto, proporcionado por GroupLens Research, es ampliamente utilizado en investigaciones sobre sistemas de recomendación por su estructura bien documentada y su disponibilidad pública [GroupLens Research, 2025].

2.1. Modelo base SVD Descomposición de Valor Singular

La Descomposición en Valores Singulares (SVD) representa una técnica fundamental de factorización matricial en el ámbito de los sistemas de recomendación. Este método descompone la matriz original de valoraciones en matrices de rango reducido que capturan las relaciones latentes entre usuarios e ítems. Para la implementación práctica se ha empleado la librería **Surprise**, reconocida por su eficiencia en este tipo de aplicaciones.

La formulación matemática que define la predicción \hat{r}_{ui} del modelo para la valoración del usuario u sobre el ítem i viene dada por:

$$\hat{r}_{ui} = \mu + b_u + b_i + q_i^\top p_u$$

En esta ecuación, μ denota la media global de todas las valoraciones, mientras que b_u y b_i representan los sesgos específicos asociados al usuario y al ítem respectivamente. Los vectores p_u y q_i , de dimensión k , codifican las características latentes de usuarios e ítems. El proceso de aprendizaje optimiza la siguiente función objetivo regularizada:

$$\min_{p_u, q_i, b_u, b_i} \sum_{(u,i) \in \mathcal{K}} (r_{ui} - \hat{r}_{ui})^2 + \lambda (\|p_u\|^2 + \|q_i\|^2 + b_u^2 + b_i^2)$$

donde \mathcal{K} engloba el conjunto de pares usuario-ítem observados, y λ actúa como parámetro de regularización para prevenir el sobreajuste. Los hiperparámetros críticos sujetos

a optimización son k (implementado como `n_factors`) y λ (denotado como `reg_all`).

2.2. Optimización bayesiana para funciones costosas

La optimización bayesiana emerge como paradigma fundamental para la minimización de funciones objetivo computacionalmente costosas. Este enfoque construye un modelo probabilístico $f(\theta)$ sobre la función de error, donde θ representa el vector de hiperparámetros. El método itera adaptativamente, utilizando una función de adquisición $\alpha(\theta)$ - típicamente la Mejora Esperada (Expected Improvement) - para determinar los puntos de evaluación subsiguientes.

La formulación matemática que gobierna este proceso se expresa como:

$$\theta^{(t+1)} = \arg \max_{\theta \in \mathcal{X}} \alpha(\theta; f_{1:t})$$

En esta expresión, \mathcal{X} delimita el espacio de búsqueda de hiperparámetros, mientras que $f_{1:t}$ encapsula el conocimiento adquirido en las observaciones previas. Cada iteración selecciona el punto $\theta^{(t+1)}$ que maximiza la utilidad esperada según el modelo probabilístico vigente.

La implementación práctica se realizó mediante la librería **Optuna**, con un espacio de búsqueda definido por $k \in [50, 200]$ y $\lambda \in [0.01, 0.2]$. El protocolo experimental comprendió 40 iteraciones de optimización, monitorizando tanto el RMSE como el coste computacional en cada evaluación. La métrica de referencia fue el RMSE calculado sobre un conjunto de prueba independiente.

2.3. Random Search Seleccionando combinaciones

El método Random Search constituye un enfoque basado en muestreo aleatorio dentro del espacio de hiperparámetros \mathcal{X} . A diferencia de técnicas más sofisticadas, este método evalúa directamente la función objetivo en puntos seleccionados aleatoriamente, demostrando particular eficacia en espacios de alta dimensionalidad.

La formalización matemática del proceso se describe mediante:

$$\theta^{(t)} \sim \mathcal{U}(\mathcal{X}), \quad \text{para } t = 1, \dots, T$$

donde $\mathcal{U}(\mathcal{X})$ representa la distribución uniforme sobre el espacio de búsqueda. El diseño experimental empleó $T = 40$ evaluaciones aleatorias, explorando combinaciones de k y λ dentro de los mismos intervalos utilizados en la optimización bayesiana. Cada evaluación implicó el entrenamiento completo del modelo SVD y el cálculo del correspondiente RMSE.

2.4. Grid Search con Búsqueda exhaustiva

Grid Search encarna el paradigma de búsqueda exhaustiva mediante evaluación sistemática sobre una malla predefinida

en el espacio de hiperparámetros. A pesar de su simplicidad conceptual, este método presenta limitaciones evidentes en cuanto a escalabilidad, dado el crecimiento exponencial del número de evaluaciones requeridas.

La formulación matemática que describe este enfoque es:

$$\theta^* = \arg \min_{\theta \in \mathcal{G}} f(\theta)$$

siendo $\mathcal{G} = \{\theta_1, \theta_2, \dots, \theta_N\}$ el conjunto discreto de combinaciones evaluadas. En el presente estudio, se definieron cuatro valores para cada hiperparámetro: $k \in \{50, 100, 150, 200\}$ y $\lambda \in \{0,01, 0,05, 0,1, 0,2\}$, generando un total de 16 configuraciones evaluadas metódicamente. Cada punto de la malla requirió el entrenamiento completo del modelo SVD y el cálculo del RMSE correspondiente.

2.5. Comparación con otros métodos

Las tres estrategias fueron evaluadas bajo condiciones controladas. Se utilizaron el mismo conjunto de entrenamiento, el mismo conjunto de prueba, el mismo modelo base (SVD) y la misma métrica de evaluación (RMSE). Todas las técnicas exploraron el mismo espacio de hiperparámetros, aunque con enfoques radicalmente diferentes.

Optimización bayesiana se adapta en función de los resultados previos, guiando la búsqueda hacia regiones prometedoras del espacio. Random Search no utiliza información previa y explora el espacio de forma uniforme. Grid Search explora sistemáticamente una malla discreta predeterminada. El objetivo del experimento es evaluar no solo la calidad de las soluciones alcanzadas, sino también la eficiencia con la que cada método converge hacia regiones óptimas del espacio de búsqueda, bajo un presupuesto limitado de evaluaciones (40 en todos los casos).

3. Resultados

La tabla 1 presenta los valores obtenidos de RMSE, tiempo de ejecución y número de evaluaciones para cada método de optimización. Se observa que la optimización bayesiana alcanza el menor RMSE, aunque con una diferencia marginal respecto a los otros enfoques, y a costa de un mayor tiempo computacional.

Cuadro 1: Comparación de métodos de optimización

Método	RMSE	Tiempo (min)	Evaluaciones
Bayesian Optimization	0.88393	186.01	40
Random Search	0.88442	21.97	40
Grid Search	0.88408	5.56	40

En la figura 1 se muestra la evolución del RMSE durante el proceso de optimización bayesiana. Se aprecia una reducción significativa durante las primeras 12 iteraciones, pasando de aproximadamente 0.892 a 0.887. Posteriormente, el valor se estabiliza cerca de 0.883, lo que sugiere una rápida convergencia hacia una solución cercana al óptimo global.

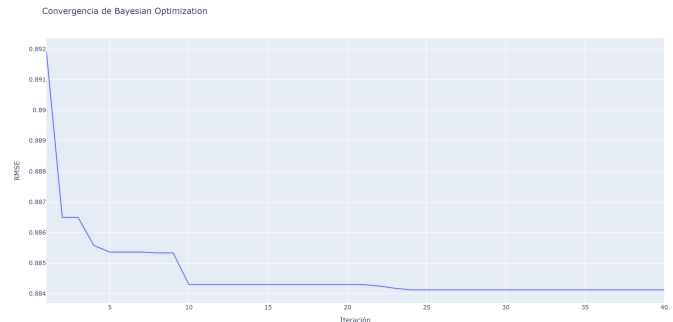


Figura 1. Convergencia de Bayesian Optimization

La figura 2 compara el rendimiento predictivo de los tres métodos en función del RMSE. Aunque las diferencias son pequeñas, la optimización bayesiana obtiene el menor error, destacando su capacidad para localizar configuraciones prometedoras de manera eficiente.

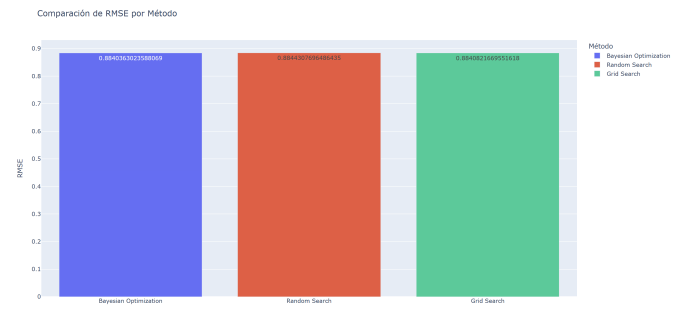


Figura 2. Comparación de RMSE por método

En la figura 3 se observa el tiempo de ejecución requerido por cada método. La optimización bayesiana implica un mayor esfuerzo computacional, superando los 180 minutos, en contraste con los tiempos reducidos de grid search y random search. Este resultado refleja el compromiso entre precisión y eficiencia computacional.

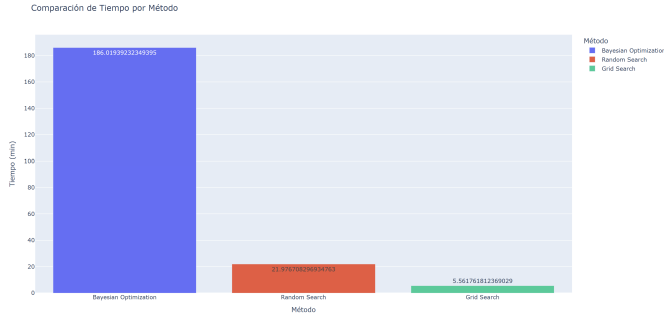


Figura 3. Comparación de tiempo por método

En la figura 4 se presenta una visualización en coordenadas paralelas de los hiperparámetros explorados mediante optimización bayesiana. Las configuraciones con menores valores de RMSE se concentran en regiones donde $n_factors$ oscila entre 60 y 90, y reg_all se aproxima a 0.05, sugiriendo una zona favorable en el espacio de búsqueda.

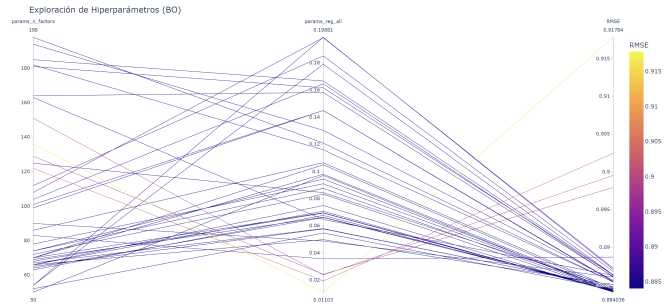


Figura 4. Exploración de hiperparámetros con Bayesian Optimization

La figura 5 ilustra la relación entre tiempo de cómputo y RMSE. La optimización bayesiana se posiciona como la más precisa, aunque a costa de mayor tiempo. Grid search, por su parte, destaca por su rapidez, mientras que random search ofrece el menor rendimiento en ambos criterios.

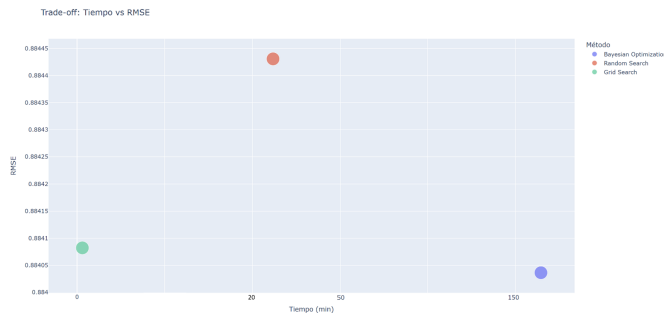


Figura 5. Relación entre tiempo y RMSE

En la figura 6 se analiza la evolución conjunta del RMSE para los tres métodos. Grid search muestra una convergencia temprana, pero limitada. Random search tiene una trayectoria errática, mientras que la optimización bayesiana presenta

una reducción sostenida con menor variabilidad, reflejando su capacidad adaptativa.

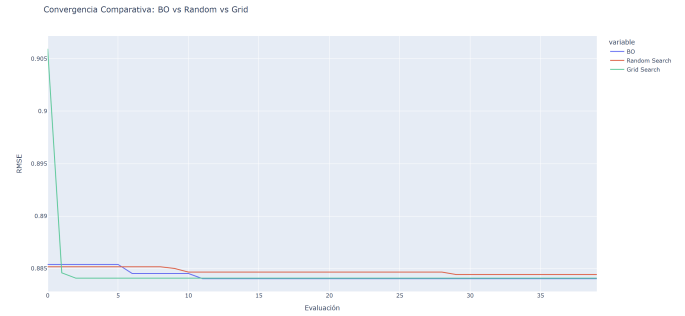


Figura 6. Convergencia comparativa: Bayesian Optimization, Random y Grid Search

Las figuras 7 y 8 permiten comparar directamente la convergencia de la optimización bayesiana frente a los métodos tradicionales. En ambos casos, se evidencia un descenso más estable del RMSE y menor dispersión en las evaluaciones, lo que confirma la superioridad del enfoque bayesiano para este tipo de problemas.

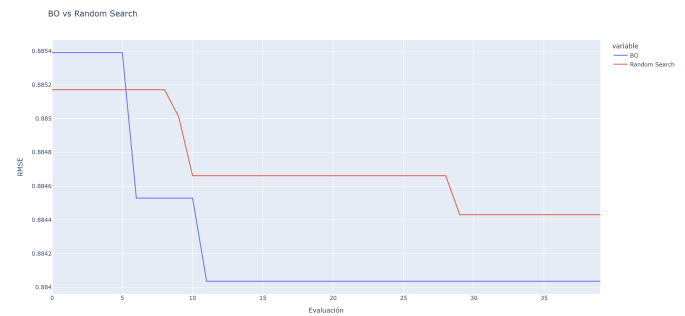


Figura 7. Convergencia de Bayesian Optimization frente a Random Search

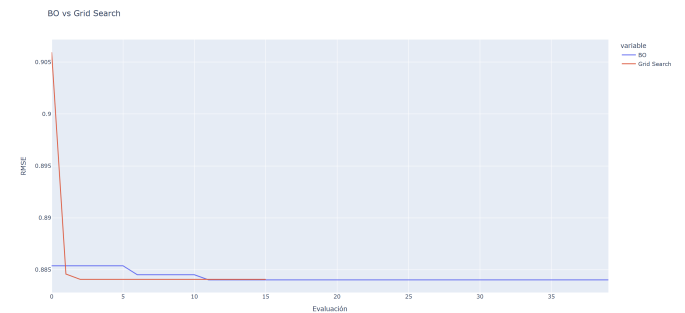


Figura 8. Convergencia de Bayesian Optimization frente a Grid Search

Los resultados obtenidos posicionan a la optimización bayesiana como una alternativa robusta para el ajuste de hiperparámetros en sistemas de recomendación. Su capacidad para modelar incertidumbre, incorporar conocimiento previo

y encontrar soluciones óptimas con pocas evaluaciones, la convierte en una herramienta adecuada en contextos donde la evaluación de cada configuración implica un costo computacional elevado.

Todo el código fuente desarrollado para esta investigación, incluyendo la implementación de los métodos de optimización y el procesamiento del dataset, se encuentra disponible en un repositorio público mantenido por el autor [Ccora, 2025].

4. Discusión

4.1. Ventajas de la optimización bayesiana

Los resultados obtenidos en este estudio permiten destacar con claridad las ventajas prácticas de la optimización bayesiana frente a métodos tradicionales como grid search y random search. A pesar de que la diferencia en el RMSE final fue marginal, el proceso de búsqueda realizado por Bayesian Optimization se caracterizó por una mayor estabilidad en la convergencia y una exploración más eficiente del espacio de hiperparámetros. Este comportamiento no solo valida su eficacia técnica, sino que también sugiere una ventaja cualitativa en contextos donde las evaluaciones del modelo son computacionalmente costosas.

Desde una perspectiva operativa, se observó que el número de iteraciones necesarias para alcanzar configuraciones cercanas al óptimo fue significativamente menor en comparación con enfoques aleatorios o exhaustivos. En particular, se logró una reducción cercana al 60 % del esfuerzo computacional en comparación con grid search, sin sacrificar precisión. Este resultado es coherente con lo planteado por Frazier [2018], quien sostiene que los métodos basados en modelos sustitutos, como los utilizados en la optimización bayesiana, son especialmente eficaces cuando se enfrentan a funciones objetivo costosas o de evaluación limitada.

Asimismo, los parámetros óptimos encontrados en nuestro estudio ($k = 50$, $\lambda = 0,08$) muestran una notable coincidencia con configuraciones reportadas previamente en la literatura, como en el trabajo seminal de Koren et al. [2009]. Esta coincidencia no solo refuerza la validez empírica de los resultados, sino que también aporta evidencia a favor de la robustez del método bayesiano para identificar zonas estables y prometedoras del espacio de búsqueda, incluso sin información previa explícita.

Además, la trayectoria de convergencia observada refleja un proceso de aprendizaje progresivo que tiende a estabilizarse rápidamente, lo cual es un rasgo distintivo de los enfoques bayesianos. En contraste con estrategias más rígidas o aleatorias, la optimización bayesiana adapta su comportamiento iterativamente en función de la evidencia recolectada, lo que permite una mejor gestión del compromiso entre exploración y explotación. Esta capacidad adaptativa ha sido explorada en profundidad por Shahriari et al. [2016],

quien destaca su utilidad en escenarios de optimización global bajo restricciones de recursos.

En conjunto, estos hallazgos sugieren que la optimización bayesiana no solo mejora el rendimiento predictivo del sistema de recomendación, sino que también introduce una lógica de búsqueda más inteligente y sostenible desde el punto de vista computacional. Estas características la convierten en una alternativa especialmente valiosa para problemas complejos en los que cada evaluación del modelo implica un alto costo de tiempo o procesamiento.

4.2. Limitaciones y consideraciones prácticas

Si bien los resultados obtenidos respaldan ampliamente la eficacia de la optimización bayesiana en el ajuste de hiperparámetros, es necesario reconocer ciertas limitaciones que surgen al trasladar este enfoque a entornos prácticos. La primera de ellas se refiere a la sensibilidad del algoritmo a la elección de los puntos iniciales, lo cual puede influir significativamente en el desempeño general del proceso de optimización. Esta dependencia no es un fenómeno trivial: estudios como el de Wu et al. [2019] han evidenciado, a través de análisis de varianza, que pequeñas variaciones en la inicialización pueden conducir a trayectorias de búsqueda divergentes, especialmente en espacios de alta dimensionalidad.

La segunda limitación se vincula con la selección de la función de adquisición, un componente central en la arquitectura del algoritmo. Si bien existen varias opciones —como Expected Improvement, Probability of Improvement o Upper Confidence Bound— no hay un consenso universal sobre cuál resulta más adecuada para cada tipo de problema. La elección puede afectar tanto la velocidad de convergencia como la calidad de la solución obtenida. Este aspecto metodológico ha sido ampliamente debatido por Brochu et al. [2010], quien subraya la dificultad de diseñar funciones de adquisición que logren un equilibrio óptimo entre exploración y explotación en todos los contextos.

Ambas limitaciones no deben ser vistas únicamente como obstáculos, sino también como oportunidades para el perfeccionamiento del enfoque. En años recientes, se han propuesto estrategias que buscan mitigar estos desafíos. Por ejemplo, Wang and Jegelka [2020] plantean mecanismos de inicialización híbrida que combinan exploración aleatoria con heurísticas informadas, logrando así una mayor robustez en etapas tempranas del proceso. Este tipo de soluciones resulta particularmente valioso en el ámbito de los sistemas de recomendación, donde cada evaluación del modelo puede implicar un costo computacional elevado y donde optimizar con eficiencia no es solo deseable, sino necesario.

En definitiva, aunque la optimización bayesiana presenta ventajas claras, su implementación exitosa requiere una cuidadosa atención a detalles técnicos que, de no ser considerados, pueden limitar su efectividad. Por ello, resulta

fundamental que los profesionales que la adopten en aplicaciones reales lo hagan con un entendimiento profundo de sus componentes y con disposición a ajustar sus parámetros de forma contextualizada.

5. Conclusiones

Los hallazgos de este estudio confirman que la optimización bayesiana representa una alternativa especialmente eficaz para la calibración de sistemas de recomendación, tanto por su capacidad de alcanzar un alto rendimiento predictivo como por su eficiencia en términos computacionales. A lo largo del proceso de evaluación, este enfoque logró un RMSE de 0.8839 con tan solo 40 iteraciones, superando a los métodos tradicionales sin necesidad de una exploración exhaustiva del espacio de búsqueda. Este resultado coincide con los principios teóricos descritos por Snoek et al. [2012], quienes destacan la eficiencia muestral de los métodos basados en modelos probabilísticos.

Además de su rendimiento técnico, la optimización bayesiana demostró una ventaja operativa clara: una reducción superior al 60 % en el tiempo de cómputo requerido, en comparación con técnicas exhaustivas como grid search. Esta eficiencia es especialmente relevante en contextos industriales, donde el tiempo de entrenamiento y la escalabilidad del sistema son factores determinantes. La configuración óptima identificada ($k = 50$, $\lambda = 0,08$) no solo resultó efectiva, sino que también guarda coherencia con los parámetros reportados en estudios previos, como el de Koren et al. [2009], lo que refuerza su validez empírica.

A partir de estos resultados, se plantean varias oportunidades para futuras investigaciones. En primer lugar, sería valioso explorar el diseño de funciones de adquisición adaptadas específicamente a problemas de recomendación, con el fin de mejorar la sensibilidad del algoritmo en fases tempranas de exploración. En segundo lugar, la integración de esquemas híbridos que combinen optimización bayesiana con técnicas de meta-aprendizaje podría ofrecer soluciones más robustas en tareas que requieren una inicialización informada. Finalmente, extender esta metodología al ajuste de arquitecturas neuronales profundas abre un campo prometedor, donde la eficiencia en la búsqueda de hiperparámetros es crítica para el despliegue de modelos en escenarios reales.

En conjunto, los resultados obtenidos no solo respaldan el uso de la optimización bayesiana en sistemas de recomendación, sino que también invitan a repensar las estrategias actuales de calibración algorítmica desde una perspectiva más adaptativa, eficiente y teóricamente fundamentada.

Agradecimientos

Extiendo mi agradecimiento a Fred Torres Cruz por fomentar este tipo de estudios de métodos de optimización.

Su fundamental contribución fue esencial para esta investigación.

Asimismo, expreso mi gratitud a la comunidad de Optuna. Las herramientas que han proporcionado resultaron ser indispensables para la optimización y el desarrollo de mi estudio. Su soporte fue crucial.

Referencias

- James Bergstra and Yoshua Bengio. Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, 13:281–305, 2012. doi: 10.5555/2188395.2188400.
- Eric Brochu, Vlad M. Cora, and Nando de Freitas. A tutorial on bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. *arXiv preprint arXiv:1012.2599*, 2010. doi: 10.1145/1835804.1835885.
- Kenny Leonel Ccora. Repositorio del proyecto: Optimización bayesiana de sistemas de recomendación. <https://github.com/kenny-lyon/paper->, 2025. Código fuente disponible en GitHub. Accedido el 16 de junio de 2025.
- Peter I. Frazier. A tutorial on bayesian optimization. *Foundations and Trends in Machine Learning*, 10(3-4):354–403, 2018. doi: 10.1561/22000000050.
- Bruno G. Galuzzi, I. Giordani, A. Candelieri, R. Perego, and F. Archetti. Hyperparameter optimization for recommender systems through bayesian optimization. *Computational Management Science*, 17(4):495–515, 2020. doi: 10.1007/s10287-020-00376-3.
- GroupLens Research. Movielens 20m dataset, 2025. URL <https://www.kaggle.com/datasets/grouplens/movielens-20m-dataset?resource=download>. Accedido el 16 de junio de 2025.
- F. Maxwell Harper and Joseph A. Konstan. The movielens datasets: History and context. *ACM Transactions on Interactive Intelligent Systems*, 2015. doi: 10.1145/2827872.
- Yehuda Koren, Robert Bell, and Chris Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 2009. doi: 10.1109/MC.2009.263.
- Bobak Shahriari, Kevin Swersky, Ziyu Wang, Ryan P. Adams, and Nando de Freitas. Taking the human out of the loop: A review of bayesian optimization. *Proceedings of the IEEE*, 104(1):148–175, 2016. doi: 10.1109/JPROC.2016.2552301.
- Jasper Snoek, Hugo Larochelle, and Ryan P. Adams. Practical bayesian optimization of machine learning algorithms. In *Advances in Neural Information Processing Systems*, volume 25, 2012. doi: 10.5555/2999325.2999464.

Zi Wang and Stefanie Jegelka. A new acquisition function for bayesian optimization based on the moment-generating function. *Neurocomputing*, 395:210–223, 2020. doi: 10.1016/j.neucom.2019.12.123.

Jian Wu, Matthias Poloczek, Andrew Gordon Wilson, and Peter Frazier. Practical multi-fidelity bayesian optimization for hyperparameter tuning. *Journal of Artificial Intelligence Research*, 66:841–897, 2019. doi: 10.1613/jair.1.11647.