

Propuesta de un esquema de recomendaciones híbrido basado en reglas secuenciales

Daniel Ranchal-Parrado
Dept. Ciencias de la Computación e I.A.
Universidad de Granada
Granada, España
danielranchal@correo.ugr.es

Carlos Porcel
Dept. Ciencias de la Computación e I.A., DaSCI
Universidad de Granada
Granada, España
cporcel@decsai.ugr.es

Jesús Alcalá-Fdez
Dept. Ciencias de la Computación e I.A., DaSCI
Universidad de Granada
Granada, España
jalcala@decsai.ugr.es

Abstract—En la era de la información, la capacidad para descubrir contenido relevante se ha convertido en una necesidad crítica para los usuarios de plataformas digitales. Los sistemas de recomendaciones han evolucionado como herramientas esenciales para filtrar y personalizar la vasta cantidad de información disponible. Este trabajo introduce un sistema de recomendaciones híbrido que integra enfoques secuenciales con técnicas de filtrado colaborativo, dirigido a capturar de manera precisa la evolución de las preferencias de los usuarios a lo largo del tiempo. Mediante el uso de Apache Spark y Scala, integramos patrones secuenciales y colaborativos para generar recomendaciones personalizadas y dinámicas. Los resultados experimentales, basados en el conjunto de datos MovieLens, demostraron que el sistema propuesto mejora la precisión de las recomendaciones en comparación con los enfoques tradicionales, abordando efectivamente problemas inherentes a los sistemas de recomendaciones como el arranque en frío y la dispersión de los datos.

Index Terms—sistemas de recomendaciones, filtrado colaborativo, esquemas secuenciales, esquemas híbridos, Apache Spark

I. INTRODUCCIÓN

En la sociedad actual, donde la cantidad de datos generados excede nuestra capacidad para absorberlos, los sistemas de recomendaciones surgen como herramientas de gran ayuda para solucionar el problema de sobrecarga de información y mejorar la experiencia de los usuarios en sus interacciones con plataformas digitales. Desde el comercio electrónico [1], hasta el *streaming* [2] y las redes sociales [3], estos sistemas personalizan la interacción de los usuarios, orientándolos hacia contenidos y productos que mejor se adaptan a sus intereses y necesidades.

Los sistemas de recomendaciones clásicos tienen una gran utilidad para su aplicación en diversidad de campos. Sin embargo, estos esquemas tienen problemas cuando existe un

drift o reorientación del comportamiento del usuario a lo largo del tiempo. Por ejemplo, supongamos que un usuario se hubiera interesado en un tiempo anterior sobre series de ciencia ficción mientras que en la actualidad solo le interesan los documentales. Esa transición de ver series a documentales en un periodo de tiempo define ese cambio de comportamiento del usuario. Los métodos existentes suponen que el comportamiento siempre va a ser el mismo en cada intervalo de tiempo, pero teniendo en cuenta que vivimos en una sociedad con gustos que cambian constantemente, estos sistemas de recomendaciones por sí solos dejan de ser efectivos. Debido a estos cambios de interés de los usuarios, se ha puesto el foco de atención en los enfoques secuenciales, que permiten extraer y procesar la información temporal que disponemos de los usuarios [4]. Sin embargo, este tipo de enfoque sufre de un problema muy recurrente y comentado en la literatura, el arranque en frío. Mientras que el arranque en frío consiste en la falta de valoraciones por parte de los usuarios en los enfoques tradicionales, en el caso de los enfoques secuenciales se traduce en falta de temporalidad. Esto supone no tener transacciones del usuario en todos los intervalos temporales, por lo que sería imposible detectar si hay cambios de comportamiento del usuario y por lo tanto hacer recomendaciones de calidad. La motivación detrás del desarrollo de un sistema de recomendaciones híbrido [5], [6] basado en reglas secuenciales radica en la necesidad de superar tanto las limitaciones de los enfoques tradicionales y de los enfoques secuenciales, incorporando la dimensión temporal de las interacciones de los usuarios para modelizar las preferencias de los usuarios a lo largo del tiempo [7], [8].

II. METODOLOGÍA

Este trabajo propone un enfoque híbrido que combina la fortaleza del filtrado colaborativo y el análisis secuencial para abordar el cambio dinámico en las preferencias de los usuarios. Se trabajó a partir del conjunto de datos MovieLens [9] y se desarrolló una librería Scala de libre distribución en el

Esta publicación es parte de la Red Temática Española de Investigación en Sistemas de Recomendación RED2022-134302-T (ELIGE-IA), financiada por el Ministerio de Ciencia e Innovación, y del proyecto de I+D+i PID2022-139297OB-I00, financiado por MICIU/AEI/10.13039/501100011033 y por FEDER/UE.

entorno de desarrollo para computación distribuida de Apache Spark, que permite abordar grandes volúmenes de datos de forma eficiente. La librería incluye distintos esquemas de recomendaciones clásicos, enfoques secuenciales e híbridos, así como diversos mecanismos de particionamiento del conjunto de datos y métricas de evaluación. La librería implementada junto con manuales de usuarios, ejemplos de uso y pruebas de unidad de la implementación, están disponibles a través del repositorio público GitHub ¹.

La metodología se centra en la extracción de patrones secuenciales [10] de las interacciones de los usuarios, integrándolas utilizando una hibridación ponderada en la que se atribuye la ponderación $1 - x$ a las recomendaciones del enfoque colaborativo y de x para el enfoque secuencial, donde x es 0.4. Una vez aplicada esta ponderación con su posterior normalización, se realiza la combinación de las predicciones de ambos enfoques y escogemos los k primeros, siendo k igual a 5.

III. RESULTADOS Y ANÁLISIS

Se realizó un estudio experimental para analizar el comportamiento de distintos esquemas de recomendaciones, que se aplicaban de manera nativa o bien mediante hibridaciones, en las que en unos casos se incluye el contexto temporal y en otros no. En dicho estudio se realizó un primer paso de obtención y preprocesamiento del conjunto de datos (MovieLens), para posteriormente analizar los distintos esquemas de recomendaciones, sus parámetros de configuración así como las posibles métricas de evaluación a aplicar en cada caso, finalizando con un análisis preciso de los resultados obtenidos. Se usó la técnica de validación cruzada con 5 iteraciones para obtener más variabilidad en los resultados. Aunque se estudiaron distintos esquemas con distintas configuraciones, como algoritmo base para mostrar la bondad de la propuesta nos podemos centrar en el filtrado colaborativo basado en usuarios con un vecindario de $k = 25$ y el coeficiente de Pearson como métrica de similitud.

Los resultados obtenidos tras los experimentos realizados demuestran que la integración de enfoques secuenciales mejora significativamente la precisión de las recomendaciones en comparación con los sistemas que no consideran la temporalidad de las interacciones. En concreto, con el enfoque colaborativo hibridado con el secuencial se obtuvo un incremento en la precisión de más del 7000% respecto de la obtenida con el enfoque colaborativo puro. Este enfoque híbrido permite, pues, modelizar el comportamiento secuencial y actual del usuario, reduciendo el impacto de los problemas asociados con el arranque en frío y la dispersión de los datos. Además, permite integrar información proveniente de distintas fuentes para considerar diversos factores relacionados con el usuario y así proporcionar recomendaciones más precisas y personalizadas.

IV. CONCLUSIONES Y TRABAJOS FUTUROS

Este trabajo introduce un sistema de recomendaciones híbrido que integra modelos secuenciales con enfoques colaborativos y ofrece una solución práctica a uno de los desafíos más persistentes en el campo de los sistemas de recomendaciones: cómo adaptar de manera precisa la personalización a los intereses dinámicos de los usuarios.

Se realizó un amplio estudio experimental para analizar el comportamiento de la hibridación de diversos esquemas de recomendaciones, incluyendo o no la variable temporal y con diversos parámetros de configuración aplicables a cada caso. Los resultados obtenidos muestran una mejora bastante considerable en las recomendaciones proporcionadas con respecto a los esquemas que no incorporan el contexto temporal, por lo que su integración en plataformas de recomendaciones puede mejorar significativamente la satisfacción y experiencia de los usuarios.

Esta propuesta abre nuevas vías para la investigación, impulsando el desarrollo de algoritmos más sofisticados y la exploración de otras formas de hibridación que puedan ofrecer una personalización aún más profunda. En concreto, uno de los puntos a mejorar es la reducción del problema del arranque en frío en el enfoque secuencial, lo que va a permitir generar mejores reglas y por lo tanto, mejores recomendaciones. En este sentido, en lugar de utilizar un algoritmo clásico de extracción de reglas secuenciales, se podría poner el foco en utilizar algoritmos de extracción de reglas que permitan tener en cuenta la valoración proporcionada por los usuarios, obteniendo reglas de más calidad.

REFERENCIAS

- [1] D. Ma, Y. Wang, J. Ma, Q. Jin, Sgnr: A social graph neural network based interactive recommendation scheme for e-commerce, *Tsinghua Science and Technology* 28 (4) (2023) 786–798. doi:10.26599/TST.2022.9010050
- [2] V. Moscato, A. Picariello, G. Sperli, An emotional recommender system for music, *IEEE Intelligent Systems* 36 (5) (2021) 57–68. doi:10.1109/MIS.2020.3026000
- [3] A. B. Suhaim, J. Berri, Context-aware recommender systems for social networks: Review, challenges and opportunities, *IEEE Access* 9 (2021) 57440–57463. doi:10.1109/ACCESS.2021.3072165
- [4] Campos, P.G., Díez, F. and Cantador, I. Time-aware recommender systems: a comprehensive survey and analysis of existing evaluation protocols. *User Model User-Adap Inter* 24, 67–119 (2014).
- [5] R. Burke, Hybrid web recommender systems, in: P. Brusilovsky, A. Kobsa, W. Nejdl (Eds.), *The Adaptive Web: Methods and Strategies of Web Personalization*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2007, pp. 377–408. doi:10.1007/978-3-540-72079-9_12
- [6] P. Castells, D. Jannach, *Recommender systems: A primer* (2023). arXiv:2302.02579
- [7] S. Wang, L. Cao, Y. Wang, Q. Z. Sheng, M. A. Orgun, D. Lian, A survey on session-based recommender systems, *ACM Computing Surveys* 54 (7) (2021)
- [8] M. Quadrana, P. Cremonesi, D. Jannach, Sequence-aware recommender systems, *ACM Computing Surveys* 51 (4) (2018)
- [9] F. M. Harper and J. A. Konstan, 'The MovieLens Datasets: History and Context', *ACM Trans. Interact. Intell. Syst.*, vol. 5, no. 4, Dec. 2015.
- [10] Y. B. Cho, Y. H. Cho, and S. H. Kim, 'Mining changes in customer buying behavior for collaborative recommendations', *Expert systems with applications*, vol. 28, no. 2, pp. 359–369, 2005.

¹<https://github.com/danitico/recsys-spark>