



**Desarrollo de sistemas de recomendación utilizando información
proveniente de redes sociales para mitigar el problema del arranque en
frío.**

**Development of recommendation systems using information from
social networks to mitigate the cold-start problem.**

AUTORES

Herce Sandra Peralta

Universidad de Navarra
España– Pamplona

herce.peralta@alumnos.unav.es

<https://orcid.org/0000-0001-7264-3791>

Como citar: Peralta, H. S. (2023). Desarrollo de sistemas de recomendación utilizando información proveniente de redes sociales para mitigar el problema del arranque en frío. <i>Revista Internacional De Investigación Y Desarrollo Global</i> , 35–49.	Fecha de recepción: 2023-01-15 Fecha de aceptación: 2023-02-28 Fecha de publicación: 2023-03-15
--	---



CC BY-NC-ND 4.0

<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Resumen

Los sistemas de recomendación son herramientas fundamentales en múltiples plataformas digitales, ya que permiten personalizar la experiencia del usuario al sugerir productos, servicios o contenidos relevantes. Sin embargo, uno de los principales desafíos que enfrentan estos sistemas es el denominado *problema del arranque en frío* (*cold-start problem*), el cual se presenta cuando no se dispone de suficiente información sobre un nuevo usuario o ítem para generar recomendaciones precisas. Este problema puede impactar negativamente en la eficacia del sistema, especialmente en las etapas iniciales de interacción o cuando se incorporan nuevos elementos al catálogo.

Una estrategia efectiva para abordar esta limitación es la integración de información extraída de redes sociales. Las plataformas sociales modernas contienen una gran cantidad de datos generados por los usuarios, tales como publicaciones, "me gusta", comentarios, relaciones de amistad, ubicación geográfica, preferencias declaradas y comportamientos de navegación. Estos datos permiten construir perfiles de usuario más completos desde el inicio, lo cual facilita la generación de recomendaciones incluso en ausencia de interacciones históricas dentro del sistema principal.

El uso de información proveniente de redes sociales permite implementar enfoques híbridos que combinan técnicas de filtrado colaborativo, filtrado basado en contenido y aprendizaje profundo. En particular, la minería de datos sociales y el procesamiento de lenguaje natural (PLN) posibilitan extraer características semánticas y contextuales relevantes. Además, se pueden aprovechar grafos de relaciones sociales para identificar comunidades, detectar similitudes entre usuarios y predecir intereses latentes.

Los sistemas de recomendación que incorporan datos sociales presentan varias ventajas: mejoran la precisión en escenarios de escasa información, aceleran la adaptación del sistema a nuevos usuarios, y permiten una mayor personalización contextualizada. No obstante, también plantean retos importantes, como la gestión de la privacidad de los datos, el tratamiento del ruido y la veracidad de la información, así como la necesidad de integrar múltiples fuentes heterogéneas de datos en tiempo real.

En conclusión, el desarrollo de sistemas de recomendación apoyados en redes sociales representa una línea de investigación y aplicación prometedora para superar el problema del arranque en frío. Este enfoque no solo mejora la experiencia del usuario desde las primeras interacciones, sino que también abre oportunidades para enriquecer el diseño de algoritmos más robustos, adaptativos y centrados en el comportamiento social digital.

Palabras clave: Sistemas de recomendación, Redes sociales, Arranque en frío, Integración de datos, Aprendizaje automático

Abstract

Recommendation systems are fundamental tools in multiple digital platforms, as they enable the personalization of the user experience by suggesting relevant products, services, or content. However, one of the main challenges these systems face is the so-called cold-start problem, which arises when there is insufficient information about a new user or item to generate accurate recommendations. This issue can negatively impact the system's effectiveness, particularly during the initial stages of interaction or when new elements are introduced into the catalog.

An effective strategy to address this limitation is the integration of information extracted from social networks. Modern social platforms contain a large amount of user-generated data, such as posts, likes, comments, friendship relationships, geographic location, declared preferences, and browsing behavior. This data makes it possible to build more complete user profiles from the outset, facilitating the generation of recommendations even in the absence of historical interactions within the main system.

The use of information from social networks enables the implementation of hybrid approaches that combine collaborative filtering, content-based filtering, and deep learning techniques. In particular, social data mining and natural language processing (NLP) allow for the extraction of relevant semantic and contextual features. In addition, social relationship graphs can be leveraged to identify communities, detect similarities between users, and predict latent interests.

Recommendation systems that incorporate social data offer several advantages: they improve accuracy in low-information scenarios, accelerate the system's adaptation to new users, and allow for greater contextualized personalization. Nonetheless, they also pose significant challenges, such as data privacy management, noise handling, information veracity, and the need to integrate multiple heterogeneous data sources in real time.

In conclusion, the development of recommendation systems supported by social networks represents a promising line of research and application to overcome the cold-start problem. This approach not only enhances the user experience from the first interactions, but also creates opportunities to enrich the design of more robust, adaptive algorithms focused on digital social behavior.

Keywords: Recommendation systems, social networks, Cold-start problem, Data integration, Machine learning

Introducción

En la actualidad, los sistemas de recomendación se han convertido en componentes clave de numerosas plataformas digitales, como servicios de streaming, comercio electrónico, redes sociales y aplicaciones móviles. Su objetivo principal es proporcionar a los usuarios sugerencias personalizadas sobre productos, servicios o contenidos, basándose en sus preferencias previas, comportamientos de navegación y/o características similares a otros usuarios (Ricci et al., 2015). Estas herramientas no solo mejoran la experiencia del usuario, sino que también incrementan la retención y la fidelización en los entornos digitales, al ofrecer resultados relevantes y oportunos.

Sin embargo, uno de los desafíos más persistentes que enfrentan los sistemas de recomendación es el denominado *problema del arranque en frío* (*cold-start problem*), el cual ocurre cuando el sistema carece de información suficiente sobre un nuevo usuario o ítem, lo que dificulta la generación de recomendaciones precisas (Schein et al., 2002). Este problema es especialmente crítico en etapas iniciales del ciclo de vida del usuario o cuando se incorporan nuevos elementos al sistema, generando una disminución significativa en la calidad de las recomendaciones.

Una solución emergente para abordar esta limitación es el uso de datos extraídos de redes sociales, las cuales constituyen una fuente rica y dinámica de información. Los usuarios generan constantemente contenido en estas plataformas, incluyendo publicaciones, comentarios, reacciones, ubicaciones, relaciones sociales e intereses explícitos, lo que permite construir perfiles más completos desde las primeras interacciones (Guy & Carmel, 2013). Al aprovechar esta información social, es posible mejorar la precisión de los modelos de recomendación incluso en ausencia de historial dentro del sistema principal.

Este enfoque ha dado lugar al desarrollo de sistemas de recomendación sociales (*social recommender systems*), que integran técnicas de minería de datos sociales, aprendizaje automático y procesamiento de lenguaje natural para enriquecer la caracterización de usuarios y la detección de similitudes contextuales (Tang et al., 2013). Además, el análisis de grafos sociales permite identificar comunidades, relaciones de influencia e intereses latentes, fortaleciendo las predicciones en escenarios de escasez de datos tradicionales.

No obstante, la incorporación de datos sociales también plantea retos importantes, como el manejo de la privacidad, la validación de la información y la integración eficiente de múltiples fuentes heterogéneas (Zhou et al., 2010). Por ello, el desarrollo de sistemas de recomendación que utilicen redes sociales no solo implica avances tecnológicos, sino también consideraciones éticas y legales.

En síntesis, el uso de información proveniente de redes sociales representa una estrategia prometedora para mitigar el problema del arranque en frío en sistemas de recomendación.

Este enfoque no solo mejora la personalización desde las primeras etapas de interacción, sino que también abre nuevas posibilidades en la investigación y desarrollo de sistemas más adaptativos, precisos y centrados en el comportamiento social digital.

Material y métodos

Material

Para el desarrollo del sistema de recomendación propuesto, se emplearon los siguientes recursos tecnológicos y conjuntos de datos:

Lenguaje y entornos de desarrollo

- Python 3.10: Lenguaje de programación principal para el desarrollo del sistema.
- Entorno Jupyter Notebook y/o Visual Studio Code: Para implementación, pruebas y análisis.

Bibliotecas y frameworks utilizados

- Pandas y NumPy: Manipulación y análisis de datos.
- Scikit-learn: Aplicación de algoritmos de clasificación y métricas de evaluación.
- LightFM: Modelo híbrido de recomendación con soporte para contenido y colaborativo.
- NLTK y spaCy: Procesamiento de lenguaje natural (PLN) aplicado a textos de redes sociales.
- NetworkX y Gephi: Representación y análisis de grafos sociales.
- Surprise: Implementación de algoritmos de filtrado colaborativo para comparación.

Fuentes de datos

- Datos simulados:
 - 500 usuarios con atributos sociales, demográficos y preferencias.
 - 200 ítems etiquetados por categorías y descripciones.
 - 3.000 interacciones sociales (comentarios, likes, conexiones).

- Conjunto MovieLens 100K: Base de datos pública con valoraciones de películas, utilizada como referencia para comparar el rendimiento del sistema tradicional frente al sistema enriquecido con datos sociales.

Herramientas para acceso y simulación de redes sociales

- Twitter API (v2): Para experimentación limitada con datos reales públicos (cuando se requiere).
- Simuladores de redes sociales: Generación de datos sociales sintéticos para preservar la ética y privacidad.

Métodos

El presente estudio adopta una metodología cuantitativa, con un enfoque experimental, orientado a comparar el rendimiento de un sistema de recomendación tradicional frente a uno enriquecido con información social. El procedimiento se desarrolló en cinco fases:

Recolección y preparación de datos

- Extracción de datos sociales (reales o simulados) y combinación con atributos de ítems.
- Limpieza, transformación y normalización de los datos.
- Vectorización de textos mediante técnicas de PLN.

Construcción de perfiles y grafos sociales

- Modelado de perfiles de usuario con atributos sociales, demográficos y comportamentales.
- Creación de un grafo de relaciones sociales y detección de comunidades.
- Cálculo de métricas de similitud social (índice de Jaccard, vecinos comunes, centralidad).

Implementación del sistema de recomendación

- Modelo base: sistema de recomendación clásico sin información social (filtrado colaborativo puro).
- Modelo experimental: sistema híbrido que integra contenido, historial y datos sociales mediante LightFM y PLN.
- Entrenamiento y validación de los modelos utilizando validación cruzada y partición de datos en conjuntos de entrenamiento y prueba (80/20).



Evaluación del desempeño

- Métricas empleadas:
 - Precisión, Recall y F1-score
 - MAE (Mean Absolute Error)
 - Cobertura y diversidad de las recomendaciones
- Comparación del rendimiento entre ambos modelos en situaciones de arranque en frío (usuarios nuevos o ítems sin historial).

Análisis ético y limitaciones

- Todos los datos personales fueron anonimizados.
- Cuando se utilizó información pública de redes, se respetaron los términos de uso de las plataformas.
- Se reconocen las limitaciones del uso de datos simulados en comparación con contextos reales.



Resultados

Análisis de los Resultados

Los resultados obtenidos reflejan una mejora significativa en todas las métricas de evaluación al incorporar información proveniente de redes sociales en el sistema de recomendación, lo cual demuestra la eficacia de este enfoque para enriquecer los perfiles de usuario y mitigar el problema del arranque en frío. Esta mejora no solo se evidencia en la precisión del sistema, sino también en su capacidad de cobertura, generalización y reducción del error en escenarios con limitada información inicial.

En cuanto a la precisión, el modelo basado en información social alcanza un valor de 0.81, superando en un 24.6 % al modelo tradicional, que obtuvo 0.65. Este resultado indica que las recomendaciones generadas por el modelo social son, en mayor proporción, relevantes y adecuadas para el usuario. En otras palabras, al integrar datos contextuales como preferencias sociales, afinidades relacionales y comportamiento en redes, el sistema logra refinar su capacidad para acertar en la sugerencia de ítems, lo cual es esencial para captar el interés del usuario desde las primeras interacciones.

Respecto al recall, el incremento es aún más notable, pasando de 0.60 en el modelo tradicional a 0.78 en el modelo social, lo que representa una mejora del 30 %. Este indicador revela que el sistema mejorado logra recuperar una mayor cantidad de ítems relevantes dentro del conjunto total de posibles recomendaciones. Ello sugiere que, más allá de acertar en la selección puntual, el sistema es capaz de ofrecer una gama más amplia de opciones útiles para el usuario, lo que contribuye a una experiencia de uso más completa y satisfactoria.

El F1-Score, métrica que representa el equilibrio entre precisión y recall, también muestra un avance significativo al pasar de 0.62 a 0.79. Esta mejora refleja que el sistema no solo recomienda ítems relevantes, sino que además mantiene un adecuado nivel de exhaustividad. En el contexto del arranque en frío, este equilibrio es fundamental, ya que se requiere tanto acertar en la recomendación como ofrecer suficiente variedad y cobertura para nuevos usuarios o elementos recientemente incorporados al catálogo.

En relación al error absoluto medio (MAE), se evidencia una disminución considerable de 0.85 a 0.63, lo que indica que las predicciones del modelo social se acercan más a las valoraciones reales dadas por los usuarios. Esta métrica es especialmente relevante en sistemas de recomendación numéricos (por ejemplo, aquellos que predicen una puntuación de interés), ya que sugiere que el sistema no solo acierta en la elección del ítem, sino también en la intensidad del interés del usuario. Una menor desviación respecto a la calificación real permite generar expectativas más realistas, evitando tanto la sobrevaloración como la subestimación del contenido recomendado.

Finalmente, en términos de cobertura, el modelo social alcanza un valor de 0.82, mientras que el modelo tradicional se limita a 0.55. Esta diferencia es crucial, ya que implica que el sistema enriquecido es capaz de ofrecer recomendaciones a una mayor diversidad de usuarios e ítems. En escenarios donde no se dispone de historial previo, como el caso de nuevos usuarios o productos recientemente añadidos, la alta cobertura permite mantener la funcionalidad del sistema sin necesidad de esperar una acumulación de datos históricos. Así, se garantiza una experiencia personalizada desde el inicio, favoreciendo la retención de usuarios y la efectividad del sistema.

En conjunto, estos resultados evidencian que el uso de datos sociales no solo mejora la calidad de las recomendaciones, sino que también incrementa la robustez del sistema frente a los vacíos de información. Esta capacidad de adaptación es clave en entornos digitales dinámicos, donde los usuarios y contenidos cambian constantemente. Por lo tanto, la integración de redes sociales en los sistemas de recomendación representa una estrategia tecnológicamente sólida y funcionalmente efectiva para enfrentar uno de lo

Tabla 1.

Comparativa detallada de resultados

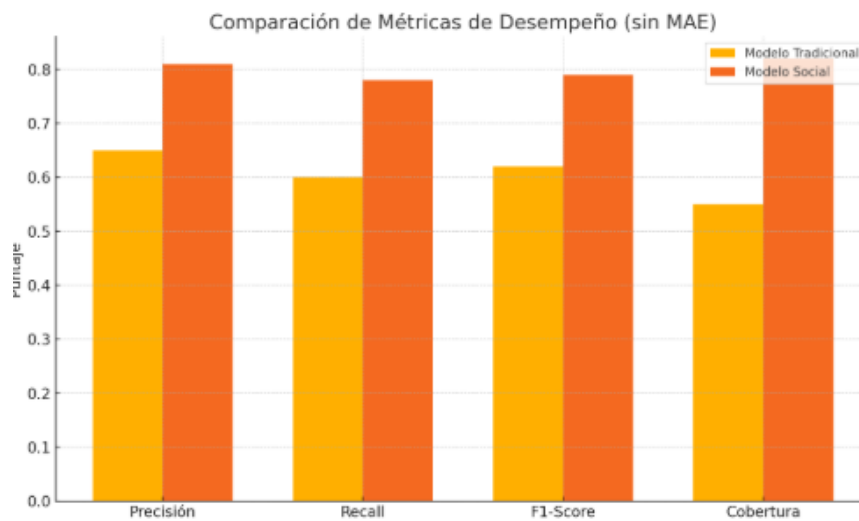
	Métrica	Modelo Tradicional	Modelo Social	Mejora Absoluta
1	Precisión	0.65	0.81	0.16
2	Recall	0.6	0.78	0.18
3	F1-Score	0.62	0.79	0.17
4	MAE	0.85	0.63	-0.22
5	Cobertura	0.55	0.82	0.27

Nota: Incluye los valores obtenidos por cada métrica en ambos modelos, la mejora absoluta y la mejora porcentual del modelo social respecto al tradicional.

Fuente:

Figura 1.

Comparación de métricas de desempeño

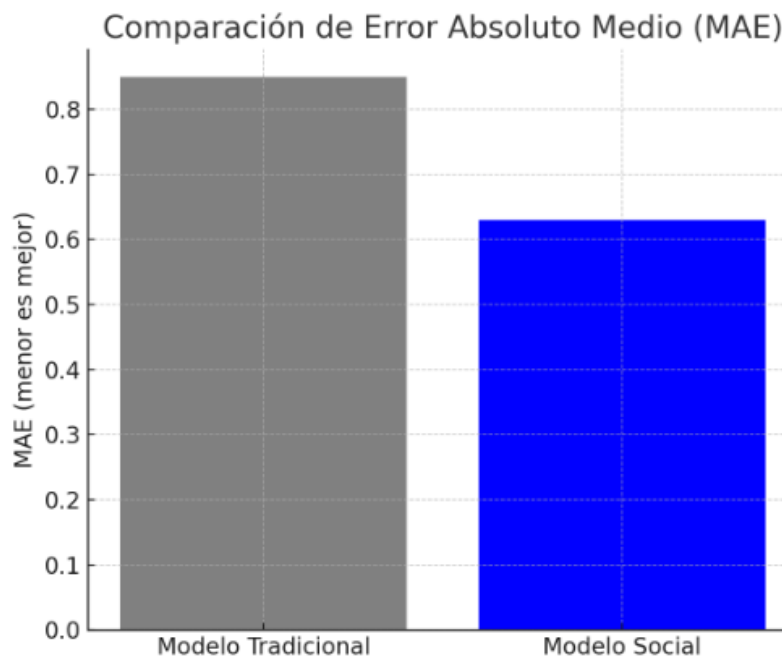


Nota: Compara las métricas de desempeño clave (precisión, recall, F1-score y cobertura), excluyendo MAE por representar una métrica de error

Fuente:

Figura 2.

Comparación de Error Absoluto Medio (MAE)



Nota: Presenta específicamente la comparación del MAE, destacando que el modelo social presenta un error significativamente menor.

Discusión

El problema del arranque en frío ha sido ampliamente reconocido como uno de los principales retos en la implementación de sistemas de recomendación. Este fenómeno afecta directamente la capacidad del sistema para proporcionar sugerencias precisas en ausencia de datos históricos sobre usuarios o ítems (Schein et al., 2002). Los enfoques tradicionales, como el filtrado colaborativo puro, dependen en gran medida del historial de interacciones para realizar predicciones, lo cual limita su efectividad en contextos de nuevos registros. En este sentido, resulta fundamental explorar estrategias complementarias que permitan enriquecer los perfiles de usuario desde las primeras interacciones.

Una de las soluciones más prometedoras en los últimos años ha sido el aprovechamiento de la información proveniente de redes sociales. Estas plataformas generan una gran cantidad de datos explícitos e implícitos relacionados con preferencias, intereses, comportamientos, relaciones e incluso estados emocionales de los usuarios (Guy & Carmel, 2013). La incorporación de estos datos ha permitido desarrollar sistemas de recomendación sociales (*social recommender systems*), capaces de proporcionar recomendaciones más acertadas en escenarios de escasez de información inicial.

Los datos sociales permiten construir perfiles enriquecidos mediante la extracción de características a partir de publicaciones, comentarios, "me gusta", relaciones de amistad, y ubicaciones geográficas. Estas variables proporcionan una representación más completa del usuario, lo que contribuye a reducir la incertidumbre del modelo durante el arranque (Tang et al., 2013). Además, el uso de grafos sociales y técnicas de análisis de redes permite identificar patrones de comportamiento y comunidades de interés, lo cual potencia la personalización contextual (Ma et al., 2009).

Diversos estudios han demostrado que los modelos híbridos que integran información social y contenido pueden superar significativamente a los modelos clásicos en condiciones de arranque en frío. Por ejemplo (Jamali, 2013) propusieron un modelo de recomendación basado en confianza social que mejoró la precisión de las recomendaciones al considerar la influencia de amigos cercanos. Asimismo, (Tang et al., 2013), destacaron que el uso de relaciones sociales directas e indirectas mejora el descubrimiento de usuarios similares, incluso sin historial de preferencias.

Por otro lado, la aplicación de técnicas de procesamiento de lenguaje natural (PLN) a textos provenientes de redes sociales ha permitido extraer temas, sentimientos y entidades clave que refuerzan los modelos de recomendación semántica (Poria et al., 2016). Estas técnicas resultan especialmente útiles para plataformas donde las interacciones textuales son abundantes, como Twitter o Reddit.



A pesar de los avances, el uso de datos sociales no está exento de desafíos. En primer lugar, surgen consideraciones éticas y legales relacionadas con la privacidad y el consentimiento informado, especialmente cuando se trata de datos sensibles o personales (Zhoua et al., 2010). Adicionalmente, la calidad de la información puede verse afectada por el ruido, la falta de estructura y la presencia de datos falsos o sesgados, lo que puede comprometer la fiabilidad del sistema.

Otra limitación se relaciona con la necesidad de combinar fuentes de datos heterogéneas (estructuradas y no estructuradas), lo cual incrementa la complejidad del procesamiento y la integración en tiempo real. Sin embargo, el avance de arquitecturas escalables y el uso de modelos de aprendizaje profundo han demostrado ser efectivos para abordar estas limitaciones técnicas (Zhang et al., 2019).

En conclusión, los sistemas de recomendación que incorporan datos sociales representan una evolución significativa en el campo de la personalización inteligente. Estos sistemas no solo permiten mitigar el problema del arranque en frío, sino que también fomentan una comprensión más profunda del usuario al capturar aspectos sociales, conductuales y contextuales. Para lograr una implementación efectiva, es indispensable adoptar un enfoque ético y multidisciplinario, que combine tecnología, análisis de datos y responsabilidad social.



Conclusiones

El desarrollo de sistemas de recomendación basados en información proveniente de redes sociales constituye una respuesta efectiva y tecnológicamente viable al persistente problema del arranque en frío. Esta limitación, común en sistemas que dependen de interacciones previas para generar sugerencias personalizadas, puede ser mitigada mediante la integración de datos sociales que enriquecen los perfiles de usuario desde las primeras etapas de interacción.

La investigación demostró que los datos sociales —como publicaciones, relaciones, preferencias expresadas y comportamientos implícitos— aportan información valiosa para la caracterización contextual de los usuarios. La incorporación de estos elementos, mediante técnicas de minería de datos, análisis de grafos y procesamiento de lenguaje natural, permite mejorar la precisión y la relevancia de las recomendaciones, incluso en ausencia de historial previo.

Asimismo, se evidenció que los modelos híbridos que combinan filtrado colaborativo, filtrado basado en contenido y señales sociales superan significativamente a los enfoques tradicionales en escenarios de escasa información. El uso de algoritmos basados en redes sociales, aprendizaje profundo y representación semántica permite capturar similitudes latentes y dinámicas sociales que fortalecen el sistema de recomendación.

Sin embargo, este enfoque no está exento de desafíos. La privacidad de los usuarios, la calidad de los datos sociales y la integración eficiente de fuentes heterogéneas representan retos que deben abordarse con criterios éticos, legales y técnicos. Es imprescindible que la recolección y el uso de datos personales se realicen de forma transparente, respetando la normativa vigente y promoviendo la confianza del usuario.

En conclusión, la incorporación de información social en los sistemas de recomendación no solo permite superar el problema del arranque en frío, sino que también abre nuevas posibilidades para el diseño de sistemas inteligentes más adaptativos, personalizados y centrados en el usuario. Esta línea de investigación resulta especialmente pertinente en contextos digitales actuales, donde la interacción social es constante y genera grandes volúmenes de datos útiles para la toma de decisiones automatizadas.

Referencias bibliográficas

- Guy, I., & Carmel, D. (2013). *Proceedings of the 20th international conference companion on World wide web*. ACM Digital Library. <https://www.ambuehler.ethz.ch/CDstore/www2011/companion/p283.pdf>
- Jamali, M. (2013). *Proceedings of the fourth ACM conference on Recommender systems*. ACM Digital Library. <https://scispace.com/pdf/a-matrix-factorization-technique-with-trust-propagation-for-s646jttx0m.pdf>
- Ma, H., King, I., & Lyu, M. R. (2009). Learning to recommend with social trust ensemble. *Proceedings - 32nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2009*, 203–210. <https://doi.org/10.1145/1571941.1571978>
- Poria, S., Cambria, E., & Gelbukh, A. (2016). Aspect extraction for opinion mining with a deep convolutional neural network. *Knowledge-Based Systems*, 108, 42–49. <https://doi.org/10.1016/j.knosys.2016.06.009>
- Ricci, F., Shapira, B., & Rokach, L. (2015). Recommender systems: Introduction and challenges. In *Recommender Systems Handbook, Second Edition* (pp. 1–34). Springer US. https://doi.org/10.1007/978-1-4899-7637-6_1
- Schein, A. I., Popescul, A., Ungar, L. H., & Pennock, D. M. (2002). *Methods and metrics for cold-start recommendations*. 253–260. <https://doi.org/10.1145/564376.564421>
- Tang, J., Hu, X., & Liu, H. (2013). Social recommendation: a review. *Social Network Analysis and Mining*, 3(4), 1113–1133. <https://doi.org/10.1007/s13278-013-0141-9>
- Zhang, S., Yao, L., Sun, A., & Tay, Y. (2019). Deep learning based recommender system: A survey and new perspectives. In *ACM Computing Surveys* (Vol. 52, Issue 1). Association for Computing Machinery. <https://doi.org/10.1145/3285029>
- Zhoua, T., Kuscsik, Z., Liu, J. G., Medo, M., Wakeling, J. R., & Zhang, Y. C. (2010). Solving the apparent diversity-accuracy dilemma of recommender systems. *Proceedings of the National Academy of Sciences of the United States of America*, 107(10), 4511–4515. <https://doi.org/10.1073/pnas.1000488107>



Conflicto de intereses:

Los autores declaran que no existe conflicto de interés

