



Tecnológico de Monterrey

Sistemas de recomendación

Entrega 1

Equipo:

Abraham Cabanzo Jimenez A01794355

Ignacio Antonio Ruiz Guerra A00889972

Moisés Díaz Malagón A01208580

Maestría en inteligencia artificial aplicada

1. Introducción a los sistemas de recomendación

Los sistemas de recomendación son indispensables para las organizaciones, en especial para las empresas de venta de productos en línea y las de distribución de contenido multimedia (Anastasiu et al. 2016). Empresas de este tipo requieren hacer buenas recomendaciones a los clientes para incrementar su valor, incentivando el tiempo que pasan en la plataforma o las ventas que generan al sugerir productos. Los sistemas de recomendación son sistemas de software que seleccionan productos y los recomiendan a clientes específicos, utilizan históricos de consumo y datos de satisfacción (Anastasiu et al. 2016).

Adicionalmente explican Elahi, Beheshti y Goluguri (2022) que los sistemas de recomendación modernos requieren acceder y entender grandes volúmenes de datos contenidos en diferentes centros de almacenamiento, ubicados en localidades potencialmente dispersas. Algoritmos paralelos y cómputo distribuido son necesarios para lidiar con la cantidad de usuarios y productos digitales en el mercado actual (Anastasiu et al. 2016), y para ello se requiere de herramientas como MapReduce y Spark.

A modo de contextualización, a continuación se explica la importancia y relevancia de los sistemas de recomendación en dos industrias, el comercio electrónico y distribución de contenido multimedia por streaming.

Comercio electrónico:

En el mundo actual del comercio electrónico globalizado, los sistemas de recomendación se han convertido en eslabones principales para el incremento de ventas, la satisfacción del cliente y la personalización de las plataformas a los gustos y deseos de los consumidores.

- **Aumento de las ventas:** Los sistemas de recomendación pueden influir significativamente en las decisiones de compra de los clientes al presentarles productos relevantes según sus preferencias y comportamientos de compra anteriores basado en usuarios para hacer recomendaciones más certeras aumentando las ventas al sugerir productos complementarios o alternativas que el cliente podría no haber considerado inicialmente.
- **Personalización:** Al analizar el historial de navegación, búsquedas clasificando y dando rating a los ítems o productos que un usuario genera, los sistemas de recomendación pueden ofrecer recomendaciones personalizadas que se ajusten a sus intereses y necesidades específicas además de cruzar intereses con otros usuarios y dar una recomendación de posibles productos creando una experiencia de compra más personalizada, lo que puede mejorar la retención de clientes y la lealtad a la marca.
- **Mejora de la experiencia del usuario:** Al simplificar el proceso de búsqueda y selección de productos ya que no se pierde el tiempo en la búsqueda, los sistemas de recomendación mejoran la experiencia del usuario al hacerla

más conveniente y eficiente al desplegar los posibles productos. Esto puede reducir la fricción en el proceso de compra y aumentar la satisfacción del cliente.

Entre los tipos de sistemas de recomendación aplicables para la industria del comercio electrónico podemos mencionar (Elahi, Beheshti y Goluguri, 2022):

- Recomendación basada en contenido: Analiza las características de los productos y las preferencias del usuario para hacer recomendaciones personalizadas igual basada en la matriz de preferencias valorando a todos los usuarios y calificaciones.
- Filtrado colaborativo: Examina el comportamiento de compra de usuarios similares para sugerir productos siendo esta la base de Content-Based Recommendation.
- Sistemas Híbridos: Combina múltiples enfoques, como basado en contenido y filtrado colaborativo, para mejorar la calidad de las recomendaciones como se puede apreciar en la siguiente imagen.

Servicios de Streaming:

Por otro lado, la creciente popularidad de las aplicaciones de distribución de video streaming como Netflix, HBO y Disney+, ha hecho que busquen incrementar los márgenes de ganancia, incrementando la cantidad de horas que un usuario pasa en la plataforma. Para ello es imprescindible recomendar contenido que sea del gusto de los espectadores de acuerdo con aquello que ha consumido anteriormente, su perfil de gustos personales, y de contenido nuevo similar al consumido por las personas de su círculo social. Es decir, en este tipo de plataformas se hace uso de múltiples técnicas combinadas de recomendación para hacer que los usuarios descubran nuevo contenido, pasen más horas en la plataforma y tengan una mejor experiencia general.

- Descubrimiento de contenido: Con bibliotecas de contenido bastante amplias, los servicios de streaming dependen de los sistemas de recomendación para ayudar a los usuarios a descubrir películas, programas de televisión, series o música que puedan ser de su interés. Esto es crucial para mantener a los usuarios comprometidos y suscritos al servicio y la experiencia de este streaming sea más user friendly.
- Retención de usuarios: Al proporcionar recomendaciones precisas y personalizadas, los servicios de streaming pueden aumentar la retención de usuarios al mantenerlos comprometidos con contenido relevante a su interés y tener usuarios más activos para el sistema de streaming. Esto es especialmente importante en un mercado altamente competitivo donde la retención de usuarios es clave para el éxito del negocio además de tener más datos para más recomendaciones.

- Mejora de la satisfacción del usuario: Al ofrecer recomendaciones relevantes y diversificadas, los sistemas de recomendación pueden mejorar la satisfacción del usuario al proporcionar una experiencia de consumo de contenido más personalizada y significativa para los usuarios.

Entre los métodos aplicables a este tipo de plataformas podemos encontrar los siguientes (Elahi, Beheshti y Goluguri, 2022):

- Filtrado colaborativo: examinar el historial de visualización y las preferencias de usuarios similares para hacer recomendaciones.
- Recomendación basada en contenido: analizar las características del contenido y las preferencias del usuario para sugerir mejor y más adecuado contenido a los usuarios basado en los datos de los mismos usuario y recomendaciones de otros usuarios.
- Sistemas de recomendación modernos: utilizar técnicas avanzadas como el aprendizaje profundo para mejorar la precisión de las recomendaciones introduciendo el IA en las recomendaciones para tener recomendaciones más rápidas y precisas.

2. Describe con detalle y explica los diferentes tipos de sistemas de recomendación.

De acuerdo con Ruy y Dutta (2022), los sistemas de recomendación se pueden categorizar en 3 grandes tipos: basados en contenido, colaborativos e híbridos.

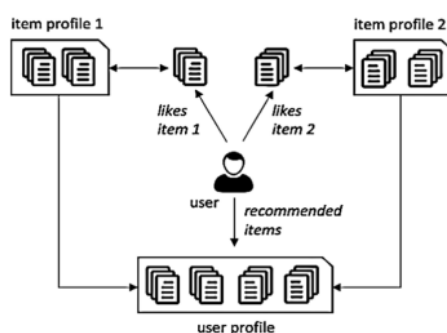


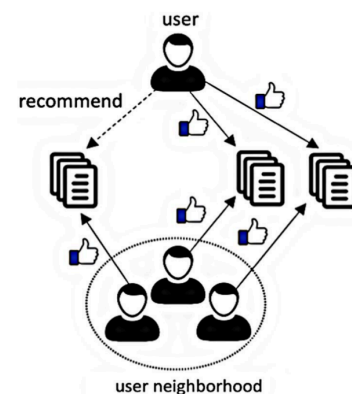
Figura 1. Sistema de recomendación basado en contenido (Roy y Dutta, 2022)

El sistema de recomendación **basado en contenido** requiere que los productos estén asociados en grupos de acuerdo con sus características, a estos grupos se les conoce como perfiles de producto. Por ejemplo, en el caso de una película, características como el director, actores o género, pueden hacer que pertenezca a un grupo determinado y que en este mismo grupo existan películas con ciertas similitudes. Cuando una persona califica como positivo un producto, los demás productos contenidos en el perfil del producto seleccionado son las recomendaciones. Si

califica como positivos múltiples productos, todos los elementos de los grupos a los que pertenecen son agregados, y estos son recomendados al cliente. A este grupo agregado de elementos se le conoce como perfil de usuario (Ruy y Dutta, 2022). Se puede entender que uno de los requisitos para que este método funcione de manera adecuada es que los productos tengan suficiente información que los describa, por ejemplo metadatos. Una ventaja es que las recomendaciones se mantienen

actualizadas a los últimos gustos del usuario, es dinámica. Otra ventaja es que cada usuario tiene un grupo de productos recomendados y es totalmente independiente de otros usuarios, lo que da seguridad y privacidad. Este sistema permite recomendar productos nuevos aunque no hayan sido consumidos por otros usuarios antes, solo se requiere que el nuevo producto esté bien descrito y que forme parte de un perfil de producto, es decir soporta el llamado cold-start. Por otro lado, Ruy y Dutta (2022) explican que el sistema de recomendación de **filtro colaborativo** utiliza una medida de similaridad entre los usuarios. A diferencia del método anterior éste tiene problemas con el llamado cold-start, y también tiene implicaciones de privacidad pues es necesario compartir datos entre usuarios, sin embargo no requiere de conocimiento de las características de los productos o elementos a recomendar, y puede generar recomendaciones nuevas más allá de los gustos actuales del usuario llevándolo a descubrir nuevos productos.

La técnica de recomendación por filtro colaborativo requiere hacer un grupo de usuarios X cuyas preferencias y gustos son similares a los de un usuario A. Los nuevos elementos o productos que les gustan a los usuarios del grupo X son recomendados al usuario A. Se dice que A es vecino de X, por lo tanto, la eficiencia de este método depende de qué tan bien el algoritmo puede encontrar los vecinos de un usuario dado.



Existen dos subtipos de filtrado colaborativo:

- **Basados en memoria:** recomienda nuevos elementos tomando en consideración las preferencias de sus vecinos. Utiliza una matriz de utilidad para la predicción. Requiere la construcción de un modelo que es una función que toma la matriz de utilidad como parámetro o entrada.

Figura 2. Sistema de recomendación de filtro colaborativo basado en usuario.

La recomendación es una función que toma como parámetros el modelo y el perfil del usuario, por lo tanto solo puede hacer recomendaciones a los usuarios contenidos en la matriz de utilidad, si se requiere hacer recomendaciones a un nuevo usuario, es necesario agregar el perfil a la matriz, y la matriz de similitud debe ser nuevamente calculada lo que hace que esta técnica sea computacionalmente pesada (Ruy y Dutta, 2022).

$$\text{Modelo} = f(\text{matriz de utilidad})$$

$$\text{Recomendación} = f(\text{Modelo}, \text{perfil de usuario}) \mid \text{perfil de usuario} \in \text{matriz de utilidad}$$

De acuerdo con Ruy y Dutta (2022) existen 2 subtipos de filtros colaborativos basados en memoria:

1. Basado en usuario: calificación de un nuevo producto para un usuario se calcula encontrando otros usuarios vecinos que previamente han calificado el producto. Si el producto ha recibido calificaciones positivas de los vecinos, el producto es recomendado al usuario (ver figura 2).

La similaridad entre usuarios se calcula utilizando la similaridad del coseno o el coeficiente de correlación de Pearson entre vectores que representan las calificaciones de dos productos de un usuario. De esta forma se puede calcular los k usuarios más similares que han calificado el artículo objetivo en la vecindad del usuario objetivo (Anastasiu et al. 2016).

2. Basado en elemento/producto: se conforma un grupo con los elementos similares que el usuario ha calificado anteriormente. Cuando hay un nuevo producto, se calcula su similitud con este grupo utilizando un promedio ponderado de todas las calificaciones de elementos similares. Si es similar al grupo con calificaciones positivas se recomienda, y si es similar para grupo con calificaciones bajas, no se recomienda.

Igual que en el caso anterior, se utiliza similitud de coseno o el coeficiente de correlación de Pearson para calcular la similitud entre objetos (Anastasiu et al. 2016).

Para este tipo de modelos existen técnicas que permiten hacer más eficiente la computación de las recomendaciones tal como la factorización de matrices, representando la matriz usuario-producto como el producto de dos matrices de factores latentes de usuario y producto. Otra técnica que se puede utilizar es descomposición de valores singulares, que permite factorizar la matriz usuario-producto en dos matrices de menor rango, una de los factores del usuario y la segunda con los factores de los productos o elementos a recomendar (Anastasiu et al. 2016).

- **Basados en modelo:** Utilizando un algoritmo de machine learning se predice la calificación del usuario para un nuevo elemento antes no calificado. Extrae características del set de datos para hacer el modelo. Es una predicción en dos pasos: construir el modelo, hacer la predicción que es una función del modelo y el perfil de usuario. Este sistema no requiere de una matriz de utilidad ni hacer computación de similitud cada que se agrega un nuevo usuario. Puede hacer recomendaciones a usuarios no presentes en el modelo. Su precisión dependerá del algoritmo y datos utilizados para

entrenar el modelo. Pueden hacer uso de técnicas de machine learning como reducción de dimensionalidad (Ruy y Dutta, 2022).

$$\text{Recomendación} = f(\text{Modelo}, \text{perfil de usuario})$$

Los **sistemas de recomendación híbridos** combinan dos o más técnicas para eliminar las limitaciones de métodos individuales como filtrado basado en contenido o el método colaborativo. Un sistema híbrido dicen Ruy y Dutta (2022) generalmente da como resultado mayor rendimiento y precisión. Algunas de las técnicas a utilizar son las siguientes:

- Meta-level: un modelo pre-entrenado se utiliza como entrada de otro sistema de recomendación.
- Combinación de características: las características de un sistema de recomendación se ingresan en otro método de recomendación.
- Incremento de características: el resultado de un modelo se aplica como entrada de un segundo.
- Hibridación mixta: la salida de diferentes sistemas de recomendación se combina.
- Hibridación en cascada: Un sistema improvisa la salida de otro.
- Hibridación cambiante: seleccionar un sistema de recomendación basado en el requerimiento del momento.
- Hibridación ponderada: Los resultados de diferentes técnicas son ponderadas con pesos para generar una única recomendación final.

Existen otros métodos como recomendaciones por contexto, que utilizan datos como tiempo, ubicación, texto u otra información asociada con los usuarios y productos. Estas técnicas son utilizadas comúnmente para hacer un filtrado previo o final de las recomendaciones realizadas (Anastasiu et al. 2016).

Finalmente, existe un problema inherente en los sistemas de recomendación llamado el problema del “cold-start”. Es común tener que agregar productos y usuarios a diario en las aplicaciones. Existen dos tipos de actualizaciones que se deben hacer cuando se agregan nuevos elementos. Actualización periódica que sucede a diario o de manera semanal cuando se re-entrena el modelo o se recalculan las matrices con todos los datos disponibles; y por otro lado, un sistema de actualización online que calcula recomendaciones de manera incremental sin re-entrenar (Anastasiu et al. 2016).

3. Identifica las metas y objetivos de los dos sistemas de recomendación de acuerdo con la información desarrollada en el punto 1.

Los objetivos de los sistemas de recomendación en el comercio electrónico se centran en aumentar las ventas, personalizar la experiencia del usuario y mejorar la satisfacción del cliente, mientras que en las plataformas de streaming, se enfocan en el descubrimiento de contenido, la retención de usuarios y la satisfacción del usuario.

En el comercio electrónico

- **Aumento de las ventas:** El objetivo principal es influir significativamente en las decisiones de compra de los clientes al presentarles productos relevantes según sus preferencias y comportamientos de compra anteriores. Esto se logra al sugerir productos complementarios o alternativas que el cliente podría no haber considerado inicialmente (Elahi, Beheshti y Goluguri, 2022).
- **Personalización:** Se busca ofrecer recomendaciones personalizadas que se ajusten a los intereses y necesidades específicas de cada usuario, utilizando su historial de navegación, búsquedas y calificaciones de productos. Esto mejora la experiencia de compra y puede mejorar la retención de clientes y la lealtad a la marca (Elahi, Beheshti y Goluguri, 2022).
- **Mejora de la experiencia del usuario:** Simplificar el proceso de búsqueda y selección de productos para hacerlo más conveniente y eficiente, reduciendo la fricción en el proceso de compra y aumentando la satisfacción del cliente (Elahi, Beheshti y Goluguri, 2022).

En plataformas de streaming

- **Descubrimiento de contenido:** Ayudar a los usuarios a descubrir películas, programas de televisión, series o música que puedan ser de su interés en una biblioteca de contenido amplia. Esto es crucial para mantener a los usuarios comprometidos y suscritos al servicio (Elahi, Beheshti y Goluguri, 2022).
- **Retención de usuarios:** Incrementar la retención de usuarios al mantenerlos comprometidos con contenido relevante a sus intereses, lo que resulta en usuarios más activos para la plataforma de streaming. Esto es especialmente importante en un mercado altamente competitivo (Elahi, Beheshti y Goluguri, 2022).
- **Mejora de la satisfacción del usuario:** Ofrecer recomendaciones relevantes y diversificadas para proporcionar una experiencia de consumo de contenido más personalizada y significativa para los usuarios, lo que aumenta su satisfacción general con el servicio (Elahi, Beheshti y Goluguri, 2022).

4. Crea un repositorio GitHub privado por equipo con la información de los integrantes del equipo.

<https://github.com/moisesdiazm/sistema-recomendacion-bigdata-mna>

Referencias:

- Anastasiu, D. C., Christakopoulou, E., Smith, S., Sharma, M., y Karypis, G. (2016). *Big data and recommender systems*. Recuperado de: https://www.researchgate.net/publication/312042146_Big_Data_and_Recommender_Systems
- Elahi, M., Beheshti, A., y Goluguri, S. R. (2021). Recommender systems: Challenges and opportunities in the age of big data and artificial intelligence. *Data Science and Its Applications*, 15-39. Recuperado de: https://www.researchgate.net/publication/353039814_Recommender_Systems_Challenges_and_Opportunities_in_the_Age_of_Big_Data_and_Artificial_Intelligence
- Roy, D., y Dutta, M. (2022). A systematic review and research perspective on recommender systems. *Journal of Big Data*, 9(1), 59. Recuperado de: <https://journalofbigdata.springeropen.com/articles/10.1186/s40537-022-00592-5>