



**Tecnológico
de Monterrey**

Análisis de grandes volúmenes de datos

Sistema de Recomendación Inicio de Proyecto

Nombres y matrículas de los integrantes del equipo:

Zmaltzin Paola Godinez Juárez

A01794538

1. Introducción

Los sistemas de recomendación son algoritmos de aprendizaje máquina diseñados para sugerir productos, servicios o contenido a los usuarios en función de sus preferencias y comportamientos previos. Estos sistemas desempeñan actualmente un papel importante en la mejora de la experiencia del usuario y uno de sus principales objetivos es maximizar la satisfacción del cliente. Dos áreas que se destacan por su dependencia en los sistemas de recomendación son el comercio electrónico y el entretenimiento digital.

En el comercio electrónico, los sistemas de recomendación son fundamentales para aumentar las ventas y fomentar la fidelidad del cliente. Utilizan técnicas de filtrado colaborativo y análisis de comportamiento del usuario para ofrecer productos relevantes y personalizados, lo que mejora la experiencia de compra y facilita la toma de decisiones. Además, permiten a las empresas comprender mejor los patrones de compra y adaptar sus estrategias de marketing y ventas en consecuencia.

En la industria del entretenimiento digital, los sistemas de recomendación son esenciales para ayudar a los usuarios a descubrir contenido relevante en un vasto número de opciones. Plataformas de streaming de video, música y juegos utilizan algoritmos sofisticados para analizar el historial de visualización, escucha o juego de un usuario, para posteriormente ofrecer recomendaciones personalizadas. Esto no solo mejora la satisfacción del usuario al proporcionar contenido que se ajusta a sus gustos, sino que también ayuda a aumentar la retención de los usuarios con la plataforma.

Si bien, la importancia de los sistemas de recomendación radica en su capacidad para crear experiencias más satisfactorias y atractivas para los usuarios la motivación principal de las organizaciones y las empresas se encuentra en el ingreso asociado a esta tecnología. De acuerdo a Ma, N. J., et al. (2017), el promedio de ganancias de una empresa asociada a los sistemas de recomendación es alrededor del 2-5%, sin embargo, en algunos casos puede llegar a ser hasta alrededor del 20%. Esto considerando que aproximadamente el 15% de las personas adultas en E.U. han confirmado que han comprado un producto recomendado por el sistema.

En este trabajo se presenta una descripción de la clasificación de los sistemas de recomendación, así como los tipos de sistemas, metas y objetivos que cada sistema considera e implementa.

2. Clasificación de los sistemas de recomendación

Los sistemas de recomendación pueden ser clasificados de acuerdo a los diferentes aspectos que dirigen al sistema, debido a esto existe una variedad de clasificaciones, en este apartado se describirán las más utilizadas.

Sistemas de recomendación de acuerdo a su **dominio**. En este caso los dominios más comunes son música, compras de supermercado, películas o noticias. Estos sistemas utilizan tanto la información anterior como las búsquedas actuales para sugerir los productos dependiendo el dominio. Los sistemas de recomendación también pueden ser clasificados de acuerdo al **propósito**. Ejemplos de este tipo de clasificación serían ventas o educación. Otra forma de clasificación es el **contexto** o lo que el usuario está haciendo en ese momento, si esta en casa, fuera con amigos o de compras. En este caso el sistema no recomendará lo mismo si el usuario está de viaje o cerca de su casa. Una forma muy conocida de clasificación son los sistemas expertos y su contraparte, los sistemas que son dirigidos por otros usuarios que son similares al usuario objetivo.

Finalmente, pueden ser clasificados de acuerdo al **nivel de personalización**. Se consideran **genéricos** cuando la recomendación es realizada para todos los usuarios. Son considerados **usuario objetivo** o *'targeted'*, cuando se realizan recomendaciones de acuerdo a alguna característica o características como la edad, ingreso, estado civil, género, etc. Se denominan **transitorios** cuando están basados en las actividades recientes, como ubicación o búsquedas, y **persistentes** cuando se basan en las preferencias de una interacción e historial más largo.

3. Tipos, metas y objetivos de los sistemas de recomendación

Los diferentes tipos de sistemas de recomendación dependen de variables principales que son las que condicionan su funcionamiento. Los principales tipos se describen a continuación:

Aprendizaje por regla de asociación (*Association rule learning*). También conocido como análisis de afinidad (*affinity analysis*). En un algoritmo no supervisado que solo trabaja con datos categóricos y es ampliamente utilizado en el análisis de compras, cuidados de la salud, extracción de texto y en el uso de las páginas web. Trabaja bien con datos de tamaño pequeño como poco productos y clientes. Es fácil de implementar y de bajo costo para obtener perspectivas y patrones de los datos.

El objetivo de esta técnica es encontrar las asociaciones de artículos que ocurren juntos más de lo que esperarías de una muestra aleatoria de todas las posibilidades. Captura estas asociaciones en términos de la siguiente regla: si es antecedente, entonces es consecuencia. Por ejemplo, si un cliente compra tortillas y aguacate entonces el cliente comprara frijoles y jalapeños.

Las métricas que se utilizan para evaluar la efectividad de la asociación son: soporte, confianza y *lift*. Sin embargo, la regla de asociación debe de ser interpretada con precaución, ya que puede traer muchas co-ocurrencias poco confiables. No implica causalidad. Se recomienda realizar la prueba de independencia Chi-square (Ma, N. J., et al, 2017).

Sistema de recomendación basados en contenido (*Content-based recommendation systems*). Esta técnica es un método que enfatiza el análisis de los atributos de un producto para generar predicciones (Nahar, et al., 2023). s. Es particularmente aplicado para recomendar fuentes, artículos web, páginas y noticias. Necesita dos importantes datos de entrada, un perfil del producto y un perfil de usuario. El primero está conformado de un numero de características del producto que se consideran útiles para establecer si a un usuario le agrada el producto o no. El perfil de usuario contiene las mismas características solo que en este caso identifican si el usuario disfrutará el producto. El perfil de usuario se crea preguntando directamente a los usuarios la información o se infiere de referencias y compras del pasado. La mayoría de los sistemas utilizan una combinación de ambas opciones. Entre los algoritmos que se utilizan comúnmente se encuentran modelos de vectores espaciales basados en palabras para el perfil del producto (como TF/IDF), y K-Nearest neighbors o Neural Networks para el perfil de usuario.

Entre las ventajas de los sistemas basados en contenido se encuentran que son sencillos de explicar, pueden recomendar nuevos artículos, encontrar sustitutos o artículos similares y que pueden modelar preferencias transitorias y persistentes, además de su independencia de otros perfiles de usuario. Entre sus principales desventajas se encuentran que los nuevos usuarios son difíciles de manejar debido a la falta de antecedentes y que los artículos requieren de etiquetas bien estructuradas que se alinean con preferencias (Ma, N. J., et al, 2017).

Sistemas de filtrado colaborativo (*Collaborative filtering*). El filtrado colaborativo funciona como un mecanismo de recomendación universal para diverso contenido como películas o música (Nahar, N. J., et al., 2023). También es conocido como filtrado colaborativo KNN, este sistema recomienda nuevos artículos o productos basados en acciones, comportamientos o preferencias de otros usuarios (vecino más cercano o *nearest neighbors*) quienes tienen un perfil similar al usuario actual. La similitud es calculada entre usuarios para encontrar los vecinos K más similares al usuario objetivo. El siguiente paso es evaluar los artículos a ser recomendados de acuerdo a las opiniones de los vecinos, usando como peso la similitud entre el usuario objetivo y los vecinos más cercanos. Como medidas de similitud se emplean distancia Euclidiana, correlación de Pearson, correlación de rango Spearman y coseno de similitud para usuarios más difíciles.

Entre las ventajas que este método presenta se encuentran que es intuitivo, no se basa en un entendimiento de los artículos y no requiere de un entrenamiento considerable. Entre las desventajas se encuentran que requiere de una cantidad inicial de usuarios significativa para crear una base de datos de preferencias. La introducción de nuevos artículos no es simple si usuarios similares no los han evaluado. Requiere de una definición robusta de similitud y algunas veces es dependiente del dominio. Tiene algunos problemas de escalabilidad cuando se tiene un gran número de usuarios. Para un número de usuarios grande, la matriz puede volverse muy grande y dispersa ya que algunos usuarios solamente utilizan algunos artículos.

En el caso de organizaciones grandes como Amazon donde el número de clientes supera por mucho al número de productos se requiere de un cambio de aplicación de filtrado colaborativo basado en usuarios (*user-user*) a un filtrado colaborativo basado en el artículo o producto (*item-item*). Esta aproximación muestra una mayor estabilidad, de manera que no requiere reconstruir el modelo tan seguido como en el caso de la aproximación basada en usuarios. Esta aproximación fue implementada por Amazon en 1998 cuando su base de usuarios estaba incrementándose considerablemente.

Filtrado híbrido. Esta técnica une múltiples algoritmos de recomendación con la finalidad de mejorar la eficiencia y sobreponerse a los retos específicos de utilizar un solo sistema. La sinergia entre las diferentes técnicas asegura que las limitaciones de una técnica pueden ser aligeradas por la otra y por lo tanto obtener recomendaciones más efectivas y precisas. La fusión de varios métodos en un modelo puede realizarse en varias formas: corriendo algoritmos aislados y después uniendo los resultados, integrando métodos de filtrado basados en contenido en una forma colaborativa, combinar diversas técnicas colaborativas o mediante la elaboración de un sistema integrado para unir ambas aproximaciones.

Conclusiones

Los sistemas de recomendación son herramientas que fueron desarrolladas principalmente por la industria para sugerir productos de potencial interés a los usuarios. Estos pueden tener diferentes niveles de personalización y pueden estar basados en similitudes entre usuarios o entre productos. Estos algoritmos pueden ser utilizados para una gran diversidad de campos además de la venta en línea y la sugerencia de las plataformas de streaming, también nuevos temas que podrías ser interesante, sugerencias de empleos, trending topics etc.

Bibliografía

Ma, A., Stagliano, A. and Willis G. (2017). Recommendation systems. O'Reilly Online Learning Platform: Academic/Public Library Edition. Streaming Video. InfiniteSkills.

Nahar, N. J., Vishwamitra, L.K., Sukheja, D., (2023). Collaborative Learning based Recommendation System for Content Streaming Platform using Non-Negative Matrix Factorization Clustering. In *Procedia Computer Science*. 2023 230:427-435 Language: English. DOI: 10.1016/j.procs.2023.12.098, Base de datos: ScienceDirect