**Instituto Tecnológico y de Estudios Superiores de Monterrey**

Maestría en Inteligencia Artificial Aplicada



**Análisis de Grandes Volúmenes de datos**

**2.2 Inicio de proyecto: Sistema de Recomendación**

**Alumnos:**

Matthias Sibrian Illescas A01794249

Johan Andrés Castro Gómez A01793556

Luis Miguel González Gutiérrez A01794779

**Profesor:** Dr. Nestor Velasco

28 de abril del 2024

## Introducción

Según Kim Falk, autor de "Practical Recommender Systems", los sistemas de recomendación son capaces de analizar grandes volúmenes de datos de usuarios para ofrecer recomendaciones completamente diseñadas para el usuario individual, y difieren de otras maneras de acercarse a un usuario final en que utilizan información particular del usuario para ofrecer el producto que se acerque más a sus gustos particulares (Falk, 2019). Esto es importante de entender, pues un sistema de recomendación es una herramienta potencialmente mucho más poderosa que la publicidad segmentada: en vez de basarse en perfiles generados de manera agregada, basándose en características compartidas por muchos clientes y agregadas de forma masiva, los sistemas de recomendación identifican el mejor producto a ofrecer para cada cliente, de forma individual.

La importancia de los sistemas de recomendación se extiende a industrias como las telecomunicaciones y la banca, donde la capacidad de recomendar productos y servicios adaptados a las preferencias individuales de los clientes puede mantener la competitividad de las empresas y promover la lealtad del cliente. A través de este trabajo, se explorará la relevancia y el impacto de los sistemas de recomendación, destacando los diferentes tipos de sistemas de recomendación e identificando las metas y objetivos que estos tienen en distintas industrias.

De manera general, en la industria de las telecomunicaciones, estos sistemas utilizan algoritmos avanzados para analizar los patrones de uso y las preferencias de los clientes, lo que permite ofrecer servicios personalizados y recomendaciones de productos. Por ejemplo, las compañías de telecomunicaciones pueden utilizar sistemas de recomendación para sugerir planes de datos adaptados a las necesidades de cada cliente, aumentando así la satisfacción del usuario y la retención (Chu, Langford y Schapire, 2010).

Por otro lado, en la industria bancaria, estos permiten ofrecer servicios financieros personalizados y fortalecer las relaciones con los clientes. Estos sistemas analizan datos financieros y patrones de comportamiento para ofrecer recomendaciones precisas sobre productos bancarios, como cuentas de ahorro, tarjetas de crédito o inversiones. Esto ayuda a los bancos a mejorar la fidelización del cliente y a aumentar las oportunidades de venta cruzada (Kusiak, 2001), una iniciativa común en muchas compañías que están en este sector.

De nuevo, esta nueva herramienta computacional es capaz de ser usada en varias industrias, con el fin de poder empoderar iniciativas relevantes para cada empresa con el factor común de poder personalizar mejor su acercamiento a los clientes. Sea un acercamiento para ofrecer un producto más acorde a las necesidades de los clientes (como un préstamo bancario con una tasa de interés particular, una tarjeta de crédito con beneficios exclusivos, o un plan de Internet o telefonía con servicios premium) o para buscar retener a clientes basándose en sus puntos de dolor, los sistemas de recomendación son ampliamente útiles para muchas industrias, tales como telecomunicaciones o la banca.

## Diferentes tipos de sistemas de recomendación

### Filtro Colaborativo

Los sistemas de recomendación de filtrado colaborativo se basan en las calificaciones proporcionadas por múltiples usuarios para hacer recomendaciones. El principal desafío en estos modelos es la dispersión de las matrices de calificaciones, es decir, la gran cantidad de valores sin especificar. Sin embargo, las calificaciones observadas a menudo están altamente correlacionadas entre usuarios e ítems, lo que permite inferir las calificaciones no observadas.

Existen dos enfoques principales para el filtrado colaborativo:

1. Métodos basados en memoria: Estos métodos, también conocidos como filtrado colaborativo basado en vecindad, se centran en predecir las calificaciones utilizando las de usuarios o ítems similares. Se dividen en:
   1. Filtrado colaborativo basado en usuarios: Compara a un usuario con otros usuarios similares para hacer recomendaciones. Por ejemplo, si dos personas tienden a comprar productos similares, lo que uno de ellos aún no ha comprado pero el otro sí, podría ser recomendado.
   2. Filtrado colaborativo basado en ítems: Analiza la relación entre diferentes ítems. Por ejemplo, si ciertos ítems se compran juntos con frecuencia, el sistema recomendaría a un usuario uno de esos ítems si ya ha comprado el otro.

Los métodos basados en memoria son fáciles de implementar y suelen ser intuitivos para los usuarios, pero tienen problemas cuando la matriz de calificaciones es muy dispersa. Con menos datos, puede ser difícil encontrar vecinos suficientemente similares para hacer predicciones precisas.

1. Métodos basados en modelos: Estos métodos emplean técnicas de aprendizaje automático y minería de datos para crear modelos predictivos a partir de las calificaciones observadas. Incluyen árboles de decisión, modelos basados en reglas, métodos bayesianos y modelos de factores latentes. Los modelos basados en algoritmos tienden a ser más precisos, incluso con matrices dispersas, gracias a sus técnicas de optimización y a la capacidad de generalización.

El filtro colaborativo tiene la ventaja de ser muy flexible y capaz de descubrir relaciones complejas, pero puede enfrentar problemas como el "cold start" (dificultad para hacer recomendaciones para nuevos usuarios o nuevos ítems) y la "falta de datos" (cuando no hay suficientes datos para hacer recomendaciones precisas).

### Basado en contenidos

En los sistemas de recomendación basados en contenido, las características descriptivas de los ítems se emplean para sugerir recomendaciones. Se usan etiquetas, descripciones y otros metadatos relacionados con el contenido para establecer relaciones entre elementos. Además, se pueden combinar las calificaciones de un usuario y su comportamiento de compra con la información contenida en los ítems para generar sugerencias.

Por ejemplo, si un usuario califica positivamente la película Terminator, aunque no tengamos acceso a las calificaciones de otros usuarios, se pueden usar descripciones de contenido para hacer recomendaciones. Al identificar similitudes en el género u otros atributos entre Terminator y otras películas como Alien y Predator, el sistema puede recomendar estas últimas.

En los métodos basados en contenido, las descripciones de los ítems, junto con sus calificaciones, se usan como datos de entrenamiento para crear modelos de clasificación o regresión específicos para cada usuario. El objetivo es predecir si un usuario pudiera estar interesado en un elemento cuyo comportamiento de compra o calificación es desconocido.

Ventajas de los sistemas basados en contenido:

* Recomendaciones para nuevos ítems: Al utilizar información de contenido, se pueden hacer recomendaciones incluso cuando no hay suficientes datos de calificación para un ítem específico. Esto permite identificar elementos con atributos similares a los que el usuario ya ha calificado positivamente.
* Modelos personalizados: Dado que el modelo se crea específicamente para cada usuario, es posible ajustar las recomendaciones de acuerdo con sus preferencias individuales.

Desventajas de los sistemas basados en contenido:

* Recomendaciones obvias: Al basarse en palabras clave y otros atributos descriptivos, estos sistemas pueden proporcionar recomendaciones predecibles. Si un usuario nunca ha interactuado con elementos con ciertas palabras clave, no serán recomendados, lo que puede llevar a una falta de diversidad.
* Dificultad para nuevos usuarios: Aunque estos sistemas son efectivos para nuevos ítems, no son tan eficaces para nuevos usuarios. Para crear un modelo robusto, se necesita un historial significativo de calificaciones del usuario, lo que puede ser difícil para nuevos usuarios.

### Sistemas de recomendación híbridos

Los sistemas de recomendación híbridos combinan diferentes enfoques de recomendación para aprovechar las fortalezas de cada uno y abordar sus debilidades. Los principales tipos de sistemas de recomendación tradicionales son el filtrado colaborativo, el filtrado basado en contenido y los sistemas basados en conocimiento. Los sistemas híbridos buscan integrar estos enfoques para proporcionar recomendaciones más precisas y robustas.

El uso de sistemas de recomendación híbridos se justifica por los desafíos y limitaciones de los métodos tradicionales. Por ejemplo:

* Filtrado colaborativo: Se basa en datos de usuario, lo que puede llevar a problemas de "cold start " cuando no hay suficiente información sobre un usuario o ítem.
* Filtrado basado en contenido: Utiliza atributos descriptivos, pero puede ser limitado en diversidad, ya que solo recomienda elementos con características similares a lo que el usuario ya ha visto.
* Sistemas basados en conocimiento: Se apoyan en información específica, pero pueden ser menos flexibles y requerir intervención manual para especificar requisitos.

Los sistemas híbridos pueden implementarse de diversas maneras, cada una con sus ventajas y desventajas.

* Diseño basado en ensamblaje: Combina resultados de diferentes algoritmos para obtener un resultado final más robusto. Puede ser mediante promedios ponderados o combinaciones secuenciales. Este enfoque es similar a los métodos de ensamblaje en minería de datos.
* Diseño monolítico: Aquí, el sistema de recomendación se construye como un solo algoritmo integrado que usa varios tipos de datos y técnicas. Puede requerir modificaciones a los algoritmos existentes para trabajar en conjunto. Por lo tanto, en este diseño se hace difícil identificar los componentes individuales del algoritmo de recomendación.
* Sistemas mixtos: Se presentan múltiples recomendaciones de diferentes algoritmos sin combinarlas en un solo resultado. Esto puede ser útil en situaciones donde se busca diversidad y variedad en las recomendaciones.

Los sistemas híbridos también se pueden categorizar según cómo se combinan los algoritmos o enfoques de recomendación:

* Ponderados: Combinan puntuaciones de varios sistemas para obtener un resultado unificado mediante promedios ponderados. El peso puede ser estático o dinámico.
* De conmutación: Cambian entre diferentes métodos de recomendación según el contexto o las necesidades del momento.
* En cascada: Un sistema refina las recomendaciones del anterior, ajustando el resultado final.
* Aumento de características: Usa la salida de un sistema como características para el siguiente, creando un proceso de refinamiento.
* Combinación de características: Combina características de diferentes fuentes de datos en un solo sistema de recomendación.

## Metas y objetivos de un sistema de recomendación aplicado a la industria

1. Sistema de recomendación de filtrado colaborativo

**Meta Principal:** Mejorar la precisión de las recomendaciones a los usuarios basándose en las interacciones y preferencias de un grupo amplio de usuarios.

**Objetivos específicos:**

* Utilizar la información colectiva de las calificaciones y comportamientos de compra de los usuarios para predecir qué productos o servicios podrían interesar a un usuario específico.
* Superar el problema de la dispersión en las matrices de calificaciones, haciendo inferencias de los datos disponibles para ofrecer recomendaciones precisas incluso con datos incompletos.
* Minimizar los efectos de los problemas de "cold start" y "falta de datos" mediante la identificación de patrones y correlaciones entre usuarios e ítems.
* Aumentar la satisfacción del usuario y la retención al personalizar las recomendaciones de manera efectiva y precisa.

1. Sistema de recomendación basado en contenido

**Meta Principal:** Ofrecer recomendaciones personalizadas utilizando las características descriptivas de los ítems, independientemente de las calificaciones de otros usuarios.

**Objetivos específicos:**

* Desarrollar perfiles de ítems detallados basados en descripciones, etiquetas y metadatos para hacer recomendaciones precisas a usuarios con intereses similares.
* Proporcionar recomendaciones de nuevos ítems rápidamente, aprovechando la información de contenido disponible, sin necesidad de esperar a que se acumulen calificaciones de usuarios.
* Ajustar los modelos de recomendación de manera individualizada, mejorando la relevancia de las sugerencias para cada usuario específico.
* Superar la limitación de recomendaciones obvias y la falta de diversidad al integrar métodos de aprendizaje que amplíen el alcance de las recomendaciones más allá de los intereses explícitamente indicados por el usuario.

## Repositorio GitHub

Enlace del repositorio del equipo para este curso:

<https://github.com/msibriani/bigdata_equipo25>

## Referencias

C. Aggarwal. Recommender systems. Cham: Springer In- ternational Publishing., 1, 2016.

Falk, K. (2019). Practical Recommender Systems. Manning Publications.

Kusiak, A. (2001). Data mining in design of banking services. International Journal of Production Research, 39(5), 1067-1078.

Li, L., Chu, W., Langford, J., & Schapire, R. E. (2010). A contextual-bandit approach to personalized news article recommendation. In Proceedings of the 19th international conference on World wide web (pp. 661-670). ACM.