

Universidad de La Habana
Facultad de Matemáticas y Computación

SRI-Recommendation-System

Integrantes: Bruno Jesús Pire Ricardo C311

Emrys Cruz Viera C311

Angel Daniel Alonso Guevara C311

Habana, 2024

Abstract

Este trabajo presenta un enfoque híbrido para sistemas de recomendación, combinando filtrado basado en conocimiento y filtrado colaborativo, específicamente utilizando la técnica de recomendación basada en vecinos más cercanos de *items*. El objetivo es mejorar la precisión y relevancia de las recomendaciones al aprovechar las fortalezas de ambas técnicas. En la primera fase del *pipeline*, se utiliza el filtrado basado en conocimiento para identificar items relevantes según las propiedades conocidas. En la segunda fase, se aplica el filtrado colaborativo basado en items para refinar las recomendaciones, calculando la similitud entre *items* y sugiriendo aquellos que son similares a los preferidos por el usuario. Los experimentos realizados demuestran que este enfoque híbrido supera a los métodos tradicionales en términos de rendimiento y eficiencia, proporcionando recomendaciones más precisas y relevantes.

Introducción

Los Sistemas de Recomendación son herramientas esenciales en diversas plataformas digitales, como tiendas en línea, servicios de *streaming* y redes sociales. Estos sistemas buscan anticipar las preferencias futuras de los usuarios basándose en sus comportamientos pasados y en las relaciones temporales entre estos comportamientos.

Una técnica comúnmente utilizada es el filtrado basado en conocimiento (*Knowledge-Based Filtering*), que se enfoca en las propiedades conocidas de los *items* para hacer recomendaciones. Por otro lado, el filtrado colaborativo (*Collaborative Filtering*) se basa en la similitud entre usuarios o *items*. Específicamente, el método de recomendación basado en vecinos más cercanos de *items* (Item-Based Nearest Neighbor Recommendation) calcula la similitud entre *items* para recomendar aquellos que son similares a los que el usuario ha interactuado previamente.

La hibridación de estas técnicas en un *pipeline* puede mejorar significativamente la precisión y relevancia de las recomendaciones. Al combinar las fortalezas del filtrado basado en conocimiento y el filtrado colaborativo, se puede proporcionar una experiencia de usuario más personalizada y satisfactoria.

Algunos sistemas que se usan tradicionalmente

Mediante **pesos**, donde se asigna un peso a cada técnica de recomendación según su importancia. Sin embargo, ajustar estos pesos puede ser complicado y requiere un análisis constante para mantener la precisión.

Otro método es la **conmutación**, en el que el sistema elige entre diferentes técnicas de recomendación según la situación actual. Aunque esto puede ser efectivo, determinar cuándo cambiar de técnica puede ser difícil y afectar la consistencia de las recomendaciones.

La **mezcla** es otra técnica, que muestra recomendaciones de varias técnicas al mismo tiempo. Esto puede resultar en una sobrecarga de información para el usuario, dificultando la toma de decisiones.

En el enfoque de **cascada**, una técnica de recomendación refina los resultados de otra. La calidad de la recomendación final depende en gran medida de la técnica inicial, lo que puede limitar la efectividad.

Finalmente, el **aumento de características** utiliza la salida de una técnica como entrada para otra. Aunque esto puede mejorar las recomendaciones, también aumenta la complejidad computacional y el tiempo de procesamiento. A pesar de sus ventajas, los sistemas híbridos de recomendación enfrentan problemas comunes como el arranque en frío, donde aún pueden tener dificultades cuando no hay suficiente información inicial sobre el usuario. Además, la sobrecarga de información puede abrumar al usuario y reducir la efectividad del sistema. La complejidad computacional también es un desafío, ya que la combinación de múltiples técnicas puede aumentar significativamente la carga computacional y el tiempo de procesamiento.

Metodología de Hibridación

La solución propuesta se basa en la hibridación de métodos de recomendación, seleccionados en función de sus ventajas y desventajas. La aproximación utilizada fue un enfoque de *pipeline*, donde los métodos se aplican secuencialmente para refinar progresivamente los resultados.

Recomendación Colaborativa: Este método, que se basa en ratings y cálculos matemáticos como la similitud del coseno y matrices rectificadas por la media, se utiliza inicialmente por su precisión. Sin embargo, su principal desventaja radica en la dispersión de valores debido a la consideración de grandes vecindades, lo cual puede no reflejar verdaderamente las preferencias del usuario, ya que se enfoca en aspectos cuantitativos más que cualitativos.

Método Grafo: Para reducir la dispersión y mejorar la precisión de las recomendaciones, se incorporó un método del grafo que permite hacer predicciones más cercanas a los usuarios. Este método ayuda a limitar las vecindades, enfocándose en objetos más relevantes para los usuarios.

Recomendación Basada en Conocimiento: Se añadió una segunda capa utilizando la recomendación basada en conocimiento (*Knowledge-based Recommendation*), que considera parámetros cualitativos de los productos. Este enfoque permite un filtrado más específico de los ítems, aunque presenta el inconveniente de no proporcionar un rating claro, ya que todos los ítems filtrados cumplen con los requisitos cualitativos del usuario sin priorización numérica.

Selección de Productos para Recomendación

Una vez que se han filtrado los productos iniciales, el siguiente paso en el proceso es reducir aún más la cantidad de *items* a un número manejable para el usuario, evitando abrumarlo con demasiadas opciones. Se considera óptimo proporcionar entre 5 y 10 recomendaciones, un rango que puede ajustarse según las necesidades específicas.

Para lograr esta reducción, se emplea el método de grafo:

- **Análisis de la Cantidad de Productos Filtrados:** Si el número de productos filtrados es igual o menor a la cantidad deseada de recomendaciones, se utilizan directamente esos ítems. En caso contrario, se recurre al método del grafo para seleccionar aquellos productos que potencialmente podrían gustar al usuario.
- **Selección Aleatoria de *items*:** Se seleccionan aleatoriamente los productos que coinciden en ambas listas (productos filtrados y productos potenciales del método de grafo). Si se logra la cantidad deseada de productos, estos pasan al siguiente paso de procesamiento. Si no, se completa la lista con productos adicionales seleccionados aleatoriamente del conjunto generado por el método de grafo.

Predicción y Presentación de Recomendaciones

Con la lista final de productos seleccionados, se procede a la fase de predicción, donde se asigna una calificación a cada ítem en función de las preferencias del usuario.

- **Matriz de Similitud Precalculada:** Se utiliza una matriz de similitud entre ítems, la cual se mantiene precalculada debido al alto costo computacional que representa calcularla dinámicamente cada vez. Esta matriz permite comparar la similitud entre productos en base a las valoraciones previas del usuario.
- **Generación de Predicciones:** Usando la lista de valoraciones del usuario y la matriz de similitud, se calculan las predicciones para cada ítem. Los productos se ordenan de mayor a menor en función de estas predicciones.
- **Presentación al *Usuario*:** Finalmente, las recomendaciones se presentan visualmente al usuario, ordenadas.

Para mejorar y completar nuestro trabajo sobre sistemas híbridos, se recomienda tener en cuenta los siguientes aspectos:

1. Evaluación Comparativa

- Pruebas con Datos Reales: Realizar un conjunto de pruebas con datos reales y comparar el rendimiento del sistema híbrido con otros métodos individuales (colaborativo, basado en conocimiento, etc.). Además medir aspectos como precisión, cobertura, tiempo de respuesta y satisfacción del usuario.

- Estudios de Caso: Desarrollar estudios de caso detallados donde se aplique el sistema a diferentes dominios (por ejemplo, recomendaciones de películas, productos en e-commerce, etc.) para demostrar su versatilidad y eficacia.

2. Incorporación de Métodos Adicionales

- Método Basado en Contenido: Considerar agregar un método basado en contenido que utilice características específicas de los ítems (por ejemplo, descripciones textuales, etiquetas) para mejorar aún más la precisión de las recomendaciones.

- Método de Filtrado Secuencial o Boosting: Implementar una técnica de boosting que ajuste dinámicamente la ponderación de cada método en función del tipo de ítem o usuario, lo que podría afinar aún más la calidad de las recomendaciones.

3. Optimización del Rendimiento

- Uso de Algoritmos de *Machine Learning*: Integrar algoritmos de machine learning avanzados, como deep learning o modelos de aprendizaje por refuerzo, para mejorar la capacidad del sistema de aprender y adaptarse a patrones de usuario más complejos.

- Escalabilidad: Evaluar y optimizar la escalabilidad del sistema, asegurando que funcione eficientemente con grandes volúmenes de datos y usuarios.

4. Mejora de la Interfaz de Usuario

- Feedback en Tiempo Real: Implementar un sistema de retroalimentación en tiempo real, donde los usuarios puedan calificar las recomendaciones recibidas, permitiendo que el sistema se ajuste y mejore de forma continua.

- Personalización Avanzada: Desarrollar opciones de personalización más detalladas, donde los usuarios puedan ajustar manualmente los parámetros que consideren más importantes en las recomendaciones.

Conclusiones

El informe presentó un enfoque híbrido para los sistemas de recomendación, combinando técnicas de filtrado basado en conocimiento y filtrado colaborativo. En la primera fase del pipeline, utilizó el filtrado basado en conocimiento para identificar ítems relevantes según sus propiedades conocidas. Luego, aplicó el filtrado colaborativo basado en ítems, calculando la similitud entre ellos para refinar las recomendaciones y sugerir aquellos que podrían interesar al usuario.

Además, incorporó un método de grafo para reducir la dispersión de valores y mejorar la precisión, limitando las vecindades a ítems más relevantes. También añadió una capa de recomendación basada en conocimiento, que permitió un filtrado más específico de los ítems, aunque presentó el inconveniente de no ofrecer una priorización numérica clara, ya que todos los ítems filtrados cumplían con los requisitos cualitativos del usuario.

Finalmente, el informe demostró que este enfoque híbrido superó a los métodos tradicionales en términos de rendimiento y eficiencia, proporcionando recomendaciones más precisas y relevantes para los usuarios.

Bibliografía

[1] Recommender Systems, An Introduction. (Dietman Jannach, Markus Zanker, Alexander Felfernig, Gerhard Friederich) (2011), <http://www.cambridge.org/9780521493369>