



Análisis de Grandes Volúmenes de Datos (Gpo 10)

David Nava Jiménez - A01168501

Edwin David Hernández Alejandro - A01794692

Jorge Fernando Bonilla Diaz - A01793935

Nombre del entregable:

Avance de proyecto 3: Sistema de Recomendación

Domingo 09 de Junio de 2024

Contenido.

1. Revisión de alcance y objetivos	3
2. Comparación entre algoritmos	3-7
3. Conclusión	7-8
4. Referencias	8

1. Revisión de alcance y objetivos

- Implementar un algoritmo de recomendación avanzado basado en contenido.
- Se realizaron dos algoritmos los cuales consideran:
 - Un sistema de recomendación que consiste en parámetros explícitos, es decir utilizando las calificaciones que los mismos usuarios realizan del mismo.
 - Un sistema de recomendación que consiste en parámetros implícitos que se refiere a suposiciones de preferencias de un usuario considerando historial de compra, historial de búsqueda o acciones hacia los items (comportamiento de un usuario).
- Identificar y justificar ventajas y desventajas, así como diferencias entre ambos algoritmos con una tabla comparativa.
- Conclusiones basadas en el análisis realizado y destacando la importancia y aplicación de cada uno de los algoritmos.

Los objetivos que se plantean con estos algoritmos de recomendación son los siguientes:

1. En el aspecto de procesos internos de una organización, se plantea que estos algoritmos de recomendación se utilicen para mejorar la experiencia de un usuario en una plataforma digital, página web, etc.
2. Adicionalmente, en este mismo contexto de procesos internos de una organización, dichos algoritmos mejorarán los canales de servicio con los usuarios.
3. En el aspecto del cliente, se plantea una mejora en la satisfacción del cliente, con el cual se busca generar mayor lealtad y fidelidad para con la organización.
4. Finalmente, y considerando los tres puntos anteriores, se tiene como objetivo que dicho sistema de recomendación pueda repercutir directamente en la generación de mayores ingresos y a su vez mayor rentabilidad de una organización.

2. Comparación entre algoritmos

Descripción del algoritmo de recomendación avanzado con feedback explícito.

De acuerdo con (P & A, 2024) un sistema de recomendación explícito contiene información explícita del usuario que muestra la calificación o rating establecido sobre un ítem, lo que implica si a un usuario le gusta o no le gusta dicho producto o servicio.

En el caso de nuestro proyecto, se realizó un recomendador simple de películas que toma como input principal la calificación que un usuario le proporciona a cada película y con este input se hacen futuras recomendaciones de títulos de películas. Adicionalmente, se puede ocupar el promedio bayesiano que se basa en integrar al promedio la variable de popularidad de una película, adicionando la variable de número de interacciones o vistas en el caso de una película. Al igual que nuestro siguiente algoritmo se utilizan matrices para su desarrollo.

Finalmente, se puede utilizar un modelo de KNN (k-Nearest Neighbours) para identificar la similaridad considerando métricas como Distancia euclidiana, correlación de pearson o distancia Manhattan.

Descripción del algoritmo de recomendación avanzado con feedback implícito.

De acuerdo con (P & A, 2024) un sistema de recomendación implícito contiene únicamente información implícita proporcionada por el usuario que incluye el comportamiento del usuario y sus opciones dentro de un ambiente específico.

En el caso de nuestro proyecto se realizó un recomendador conocido como filtrado colaborativo, el cual aprende la relación entre usuarios y películas, donde cada usuario es único en sus opciones elegidas. La película recomendada a un nuevo usuario depende de las sugerencias hechas a usuarios anteriores con intereses similares. Este filtrado colaborativo se basa en una matriz de calificación usuario-ítem donde cada usuario puede calificar ciertas películas que proporcionan información de cuánto le gusta o le disgusta una película.

Para el sistema de recomendación implícito nos basamos en una factorización de matrices (2 matrices, matriz usuario-factores y matriz item-factores). Ocuparemos la técnica de cuadrados alternos (ALS).

Ambos algoritmos de recomendación se basan en el factor de similaridad que depende de los términos de sentimiento humano y retroalimentación observada y que se almacena en una base de datos para utilizarlos en un futuro.

A continuación se presentan dos tablas comparativas resaltando las similitudes y diferencias primero del filtrado basado en contenido comparado con el sistema de recomendación item-item (arranque en frío), así como de los sistemas de recomendación explícitos comparado con sistemas de recomendación implícitos.

Aspecto	Filtrado Basado en Contenido (Coseno)	Recomendación Item-Item (Coseno)
Precisión	Alta precisión al inicio	Alta precisión en sistemas con datos suficientes y actualizados
Complejidad Computacional	Alta, debido al cálculo de similitudes para cada par de películas y usuarios	Alta, pero puede mejorar el rendimiento mediante técnicas de reducción dimensional
Escalabilidad	Media, ya que la complejidad aumenta con el número de películas y usuarios	Alta, debido a la capacidad de calcular y almacenar similitudes entre películas
Gestión de Grandes Volúmenes de Datos	Desafiante sin optimizaciones específicas como estructuras de datos eficientes	Mejor manejo debido a los cálculos y técnicas como matrices dispersas
Adaptabilidad a actualización/inserción de nuevos datos	Baja, requiere recalcular similitudes frecuentemente	Alta, ya que la actualización de similitudes puede ser más localizada y menos costosa computacionalmente
Implementación en Sistemas Distribuidos	Baja, debido a la necesidad realizar los cálculos de manera frecuente	Alta, debido a la posibilidad de distribuir los cálculos de similitud entre películas

1. Precisión:

- KNN Coseno: Este algoritmo tiene una alta precisión ya que recomienda películas basados en la similitud directa entre las películas que el usuario ya ha interactuado y otras películas. Sin embargo, la precisión puede disminuir si los datos de los usuarios no son suficientemente representativos.
- Item-Item Coseno: Este algoritmo mantiene una precisión alta si se tiene suficiente información de la interacción entre películas. Es más potente en sistemas donde los usuarios tienen interacciones con más películas.

2. Escalabilidad:

- KNN Coseno: La escalabilidad es limitada ya que el tiempo de procesamiento y los requerimientos de memoria aumentan linealmente con el número de películas y usuarios.
- Item-Item Coseno: Este enfoque puede escalar mejor utilizando técnicas como matrices dispersas y algoritmos distribuidos para manejar la similitud entre películas.

¿Cuándo Usar Cada Técnica?

Filtrado Basado en Contenido KNN (Coseno) es mejor si:

- El sistema tiene un número manejable de películas y usuarios.
- Se necesita de una alta personalización en las recomendaciones.
- La interpretabilidad de las recomendaciones es importante.
- Se tiene suficientes datos de usuario y además son representativos.

Sistema de Recomendación Item-Item (Coseno) es mejor si:

- El sistema cuenta con grandes volúmenes de datos.
- Se busca eficiencia y escalabilidad en la recomendación.
- La necesidad de actualización rápida es indispensable.
- Se tiene acceso a datos históricos para predecir similitudes con precisión.

S/R feedback explícito	S/R feedback implícito	S/R feedback explícito	S/R feedback implícito
Similitudes		Diferencias	
Ocupa filtrado colaborativo		Más simple de implementar al considerar como input una calificación de un usuario	Más complejo de implementar al considerar como input un comportamiento de

		un usuario
Se definen matrices como input	Puede ocupar técnicas como promedio bayesiano	
Se puede obtener un output con las recomendaciones realizadas por el algoritmo	Utiliza modelos como KNN como clasificadores	Utiliza técnica como SVD o Cuadrados alternos
	Las métricas utilizadas son similaridad coseno, distancia euclidiana, distancia manhattan o correlación de pearson	Las métricas utilizadas son precisión, recall, f-1 score, matriz de confusión etc

S/R feedback explícito		S/R feedback implícito	
Ventajas	Desventajas	Ventajas	Desventajas
Más simple de implementar	Menor capacidad de obtener insights	Mayor capacidad de obtención de nuevos insights	Más complejo de implementar
	Es probable que tenga mayor información sesgada, ya que la calificación no puede representar el gusto del usuario	Lo que se evalúa es el comportamiento del usuario, lo que puede proporcionar información más real de los gustos de la persona	Puede ser un análisis subjetivo y que puede sesgar por interpretaciones de un analista

El enlace con la evidencia del feedback implícito se encuentra en la siguiente liga:

<https://github.com/drekkel/TC4034.10/blob/main/RecommendationSystem/Notebooks/Parte%203.ipynb>

3. Conclusiones

En resumen, si se necesita un sistema altamente personalizado y los datos son manejables, el Filtrado Basado en Contenido KNN (Coseno) puede ser la mejor

opción. Sin embargo, para escenarios donde la escalabilidad, eficiencia y adaptabilidad son más importantes, el Sistema de Recomendación Item-Item (Coseno) suele ser más adecuado.

Es importante resaltar que los sistemas de recomendación en general cuentan con una problemática que son los nuevos usuarios y existen otras variables con las que no se cuentan que pueden determinar las decisiones de cada usuario, estas están relacionadas con la clase-socioeconómica, edad, género, nacionalidad y ubicación. Algunos de los sistemas de recomendación que se utilizan en la industria utilizan dos o más sistemas para potencializar la precisión de recomendación que tienen y así incluir atributos que contengan más información. Coincidimos que este tipo de sistemas han revolucionado muchas industrias que basan su modelo de negocio en datos como son Netflix, Spotify, entre otras. Esto ha permitido contar con una ventaja competitiva sobre sus competidores y diferenciarse del valor que entregan a los usuarios.

El análisis detallado y los resultados se han documentado en el repositorio GitHub del equipo, estableciendo un estándar de transparencia y reproducibilidad para el proyecto.

4. Referencias

P, D., & A, M. (2024, May 14). Book recommendation system: reviewing different

techniques and approaches. *International Journal on Digital Libraries*, 1-22.

<https://o-link-springer-com.biblioteca-ils.tec.mx/content/pdf/10.1007/s00799-024-00403-7.pdf>