

Técnicas Híbridas en Sistemas de Recomendación

Alejandro Ramirez Trueba, Ana Melissa Alonso Reina, and Alejandro Lamelas Delgado

Universidad de la Habana

Resumen Las técnicas híbridas en sistemas de recomendación integran diferentes enfoques de recomendación para mejorar la calidad y eficacia de las recomendaciones. Este documento describe la combinación de filtrado colaborativo y basado en contenido, abordando las limitaciones como la escasez de datos y el sesgo.

Keywords: Sistemas de recomendación · Filtrado colaborativo · Filtrado basado en contenido.

1. Introducción

Las técnicas híbridas en sistemas de recomendación se refieren a la integración de diferentes enfoques de recomendación, tales como el filtrado colaborativo, el filtrado basado en contenido y los métodos basados en el conocimiento, con el objetivo de mejorar la eficacia y la calidad de las recomendaciones proporcionadas a los usuarios. Estas técnicas buscan combinar las fortalezas de cada enfoque individual mientras se minimizan sus debilidades, como los problemas de arranque en frío, la escasez de datos o el sesgo.

El filtrado colaborativo, que se basa en las preferencias y comportamientos de usuarios similares, puede ser limitado cuando no existen suficientes datos históricos. Por otro lado, el filtrado basado en contenido, que se enfoca en las características de los ítems para hacer recomendaciones, puede no capturar adecuadamente las preferencias complejas de los usuarios. Los métodos basados en el conocimiento utilizan información adicional, como reglas de negocio o conocimiento experto, para mejorar las recomendaciones, pero pueden ser costosos de implementar.

Al combinar estos enfoques, los sistemas híbridos pueden ofrecer recomendaciones más precisas, diversificadas y personalizadas. Por ejemplo, una técnica híbrida podría utilizar el filtrado colaborativo para generar una lista inicial de recomendaciones, que luego se filtran o ajustan utilizando un enfoque basado en contenido o reglas de conocimiento. De esta manera, se logra una mayor robustez en las recomendaciones, mejorando la satisfacción del usuario y la efectividad del sistema en diversos contextos.

En resumen, las técnicas híbridas en sistemas de recomendación son esenciales para superar las limitaciones de los métodos individuales, proporcionando una solución más completa y eficaz en la personalización de las recomendaciones.

2. Antecedentes Sistemas de Recomendacion Híbridos

El desarrollo de los sistemas de recomendación ha evolucionado significativamente desde sus inicios, marcados por métodos simples y directos, hasta las complejas estructuras actuales que utilizan algoritmos avanzados. Originalmente, estos sistemas se centraban en técnicas de filtrado basado en contenido, donde las recomendaciones se generaban a partir de la similitud directa entre los ítems y las preferencias explícitas de los usuarios. Con el tiempo, el filtrado colaborativo tomó prominencia, aprovechando las valoraciones y comportamientos de múltiples usuarios para inferir gustos y preferencias, superando así las limitaciones de los métodos basados solo en contenido. La necesidad de abordar problemas específicos como la escasa personalización y el arranque en frío llevó al desarrollo de sistemas híbridos, que combinan múltiples enfoques de filtrado para optimizar la precisión y relevancia de las recomendaciones.

Aunque muchas aplicaciones de recomendación son de hecho híbridas, poco trabajo teórico se ha centrado en cómo hibridar algoritmos y en qué situaciones se puede esperar obtener beneficios de la hibridación. Un excelente ejemplo de la combinación de diferentes variantes de algoritmos de recomendación es la competición del Netflix Prize, en la cual cientos de estudiantes e investigadores se unieron para mejorar un sistema de recomendación de películas colaborativo, hibridando cientos de diferentes técnicas y enfoques de filtrado colaborativo para mejorar la precisión general. El artículo de Robin Burke, "Sistemas de Recomendación Híbridos: Encuesta y Experimentos", es una encuesta bien conocida sobre el espacio de diseño de diferentes algoritmos híbridos de recomendación. En él, se propone una taxonomía de diferentes clases de algoritmos de recomendación. Además, Burke investiga la recomendación basada en la utilidad, que puede considerarse un subconjunto específico de los sistemas de recomendación basados en conocimiento, ya que un esquema de utilidad puede verse como otra forma específica de codificación de conocimiento explícito de personalización.

Por ejemplo, los sistemas de recomendación demográficos realizan proposiciones colaborativas basadas en perfiles demográficos de usuarios, la información demográfica puede considerarse simplemente como un elemento adicional de conocimiento del usuario que se puede explotar para determinar usuarios similares en la web y, por lo tanto, que sea una variante de un enfoque colaborativo. Por ejemplo, en el caso de que haya pocas calificaciones de usuarios disponibles, los datos demográficos pueden usarse para iniciar un sistema de recomendación.

La segunda dimensión que caracteriza a un sistema es el diseño de hibridación: el método utilizado para combinar dos o más algoritmos. Los componentes de recomendación pueden trabajar en paralelo antes de combinar sus resultados, o dos o más sistemas de recomendación individuales pueden estar conectados en una arquitectura de canalización, en la cual la salida de un recomendador sirve como entrada para el siguiente.

3. Tecnologías y Algoritmos Avanzados

Las tecnologías y algoritmos avanzados en sistemas de recomendación han evolucionado para incorporar enfoques de inteligencia artificial y aprendizaje automático más sofisticados. Los algoritmos modernos ahora utilizan técnicas como el aprendizaje profundo y las redes neuronales para analizar grandes volúmenes de datos con una precisión notable. Estos métodos permiten a los sistemas aprender y adaptarse a las preferencias y comportamientos cambiantes de los usuarios en tiempo real, mejorando significativamente la personalización y relevancia de las recomendaciones.

Las redes neuronales convolucionales y recurrentes se han aplicado para analizar secuencias temporales de interacciones, capturando patrones complejos en los datos de usuario que métodos más tradicionales podrían pasar por alto. Además, algoritmos como la factorización de matrices y las máquinas de soporte vectorial también se emplean para descubrir estructuras subyacentes en los datos, permitiendo predicciones más precisas y robustas en entornos dinámicos.

SR Híbrido de Películas

Dataset:

Para el sistema de recomendación implementado se utilizó un dataset de Grouplens de películas que encajaba con las exigencias de nuestro proyecto (\oplus MovieLens Latest Datasets). En este se pudo acceder a los generos de cada película y a los puntajes que le dieron los usuarios a dichas películas en determinado punto en el tiempo. Al procesar este dataset, los puntajes fueron modificados para quitarle peso a los puntajes más viejos y así las recomendaciones fueran mas sensibles con los intereses más actuales del usuario.

SR:

El Sistema de Recomendación que se implementó fue uno híbrido de combinación de características que combina características colaborativas (en este caso la interacción de los usuarios con las películas) con características de contenido (en este caso los géneros de las películas). El diseño que se implementó se conoce como diseño Monolítico.

Se optó por un sistema de recomendación híbrido con un diseño monolítico debido a varias razones estratégicas y técnicas:

- Simplicidad: Un diseño monolítico permite integrar elementos de varios algoritmos de recomendación en un único algoritmo.
- Rendimiento: Al utilizar un único algoritmo, lo hace potencialmente más eficiente que otros diseños de sistemas híbridos.

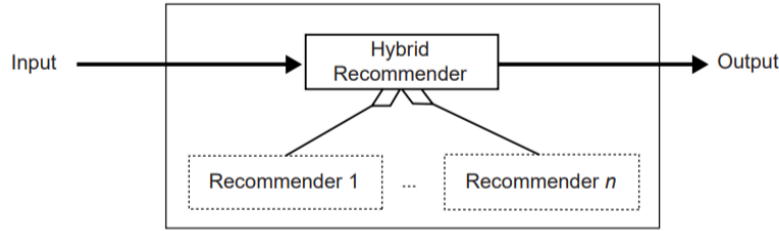


Figura 1. Diseño monolítico

- **Consistencia:** Existe menor riesgo de incompatibilidad entre los componentes por ser un único algoritmo recibiendo un input y devolviendo un output, en otros diseños como el Pipelined Hybrid (Híbrido en Canalización) si se modifica alguno de los algoritmos de recomendación que implementa, puede llegar a ser necesario modificarlos todos.

4. Análisis de las Recomendaciones

En este proyecto, se evaluó la efectividad de el sistema de recomendación mediante las métricas de precisión, recall, F1 score y Mean Average Precision (MAP), tanto a nivel de usuario individual (métricas macro) como a nivel global (métricas micro).

Precisión: Los resultados muestran que el sistema tiene un buen rendimiento en términos de precisión, con una precisión macro promedio del 73,10 % y una precisión micro global del 70.20 %. Esto indica que la mayoría de las recomendaciones realizadas por el sistema son relevantes para los usuarios, lo que demuestra la confiabilidad del sistema en la calidad de las recomendaciones que hace.

Recall: Sin embargo, las métricas de recall son más bajas, con un recall macro promedio del 47.25 % y un recall micro global del 40,77 %. Esto sugiere que, aunque las recomendaciones que se muestran son en su mayoría precisas, el sistema no está capturando una proporción suficiente de los elementos relevantes disponibles. Esto podría llevar a los usuarios a perder oportunidades de descubrir contenido que podría interesarles. Dado que el sistema está diseñado para recomendar solo 10 películas, es natural que el recall sea más bajo comparado con sistemas que intentan cubrir un espectro más amplio de posibles recomendaciones. El objetivo de este sistema es priorizar la calidad sobre la cantidad, enfocándose en las películas que más probablemente apreciarán los usuarios, en lugar de intentar capturar todas las posibles películas que podrían gustarles.

F1: El F1 score, que es una medida combinada de precisión y recall, refleja este desequilibrio, con valores de 56,25 % (macro) y 51,58 % (micro). Estos valores indican que, aunque el sistema es bueno en recomendar contenido relevante,

todavía hay margen de mejora, especialmente en la capacidad del sistema para identificar y sugerir más de los elementos que realmente interesan a los usuarios.

MAP: El Mean Average Precision (MAP) obtenido fue 76,14 %, lo que refuerza la conclusión de que el sistema no solo es capaz de identificar recomendaciones relevantes, sino que también las prioriza adecuadamente en las listas mostradas a los usuarios. Esto indica que el sistema es efectivo en proporcionar recomendaciones de alta calidad y bien ordenadas.

5. Flaquezas y Posibles Mejoras

Arranque en Frío

En una situación donde hayan pocos datos sobre los ratings (o ninguno), ocurrirá que las recomendaciones pueden llegar a ser aleatorias. Lo que hace que potencialmente se le muestre información que nada tenga que ver con el usuario al arrancar por primera vez el sistema.

Posible Solución: Disponiendo de un dataset que además contenga información demográfica de los usuarios, podría utilizarse dicha información como bootstrap para el estado inicial del sistema de recomendación.

Limitaciones del Diseño

Si bien el diseño monolítico es sencillo de implementar con una cantidad no muy extensiva de elementos, a medida que se incorporen elementos de otros sistemas de recomendación se comprobará que se va haciendo más compleja la implementación por ser un único algoritmo el que se encarga de la recomendación.

Referencias

- Recommender Systems an Introduction
- <https://link.springer.com/article/10.1023/A:1021240730564>