

Sistema de recomendación

Ernesto Rubén Palacios Gómez, 312280622
Flores González Luis Brandon, 312218342
Luis Enrique Cortez Flores, 311209219
Peto Gutierrez Emmanuel, 414008117

1. Introducción

A lo largo de nuestro día a día vamos escuchando y haciendo recomendaciones; desde que comer, ver, escuchar, visitar o hacer, que hace que nuestras experiencias y conocimientos se vayan ampliando sin darnos cuenta. Es una parte natural del ser humano hacer recomendaciones para que otros tengan las mismas experiencias, sentimientos o conocimientos que creemos debemos de compartir con los demás. No siempre las recomendaciones son hechas de seres humanos para otros seres humanos. Y para eso existen los sistemas de recomendación que nos dan recomendaciones a partir de datos que saben de nosotros, como las cosas que ya hemos visto, hecho y escuchado para que nos recomienden cosas similares que podrían llegar a interesarnos.

2. Definición del problema

Hay diferentes maneras en las que podemos identificar el uso de sistemas de recomendación, pero una manera clara y sencilla de hacerlo es por medio de una plataforma que indique una nueva película a un usuario, dado sus gustos previos. Además para lograr hacer esta plataforma se utiliza una base de conocimientos.

Esta plataforma se puede entender mejor cómo el funcionamiento que hace Netflix o Youtube para indicar nuevos artículos a sus usuarios, esta similitud sólo se hace para entender mejor el problema a resolver.

3. Descripción de la propuesta e implementación

a. Una breve descripción del objetivo de sistema y por qué es necesario un sistema de recomendación.

Para resolver este problema planteado se desarrollara un sistema de recomendación colaborativo para que este escoja una película dependiendo los gustos de un usuario dado y la información que proporcionó. Se requiere que este tenga un patrón definido como conocimiento para poder tomar una decisión y se le pueda recomendar una película tomando como base las similitudes que tiene con otros usuarios dentro de la base de conocimientos, es decir, se recomendará una película con base en el conocimiento que se tiene con los otros usuarios y lo que les gusta.

Un sistema de recomendación es necesario para las empresas porque así pueden ofrecerte productos que probablemente te gusten (películas, música, libros, anuncios, etc) y esto les permite aumentar sus ganancias.

b. Las características que considera el perfil de tu usuario.

Para la implementación del usuario solo se solicita un número (userId) y que califique del 0 al 5 un conjunto de películas que ya pertenecen a la base de datos. Además se asume que el userId es un atributo de autoincremento.

c. Las características que tu producto que se deben incluir.

Los atributos del producto, que en este caso es una película, son:

- **movieid** que es un número entero y representa una clave para la película.
- **title** que es el título de la película.
- **genres** que es el conjunto de géneros de la película.

d. Qué técnica de filtrado de información propones que se use, indicando qué tipo de algoritmo de recomendación consideras que funcionaría mejor (basado en usuarios o ítems) y por qué.

Técnica de filtrado colaborativo. Esto es por la poca información que proporciona el ítem (la película) que solo es el título y el género. Los usuarios proporcionan más información sobre este con calificaciones.

El algoritmo usado es basado en ítems. En este caso, porque vamos a buscar las películas que se parezcan más a las que el usuario les ha dado una calificación alta y así probablemente le gusten.

e. Describe mediante un ejemplo el funcionamiento de tu sistema de recomendación.

Supongamos que se quiere hacer una recomendación personalizada para el usuario cuyo id es 3. El usuario debe elegir la cantidad de películas que se le van a recomendar, para este ejemplo digamos que son 10. Primero, el usuario 3 debe calificar un número determinado de películas (en nuestro conjunto de datos califica 20). Supongamos que el conjunto de películas calificadas por 3 es C y el conjunto de películas no calificadas es NC.

Para cada i en NC se aplica el siguiente algoritmo:

- Para cada j en C se calcula la correlación de Pearson entre i y j, y se guardan los resultados.
- Se ordena el conjunto C de mayor a menor con base en su similitud con i.
- Se eligen las 5 películas del conjunto C que tienen mayor similitud a i.
- Se estima la calificación que el usuario 3 le dará a la película i, usando promedio por pesos, y se guarda ese valor.

Una vez que se tienen todas las calificaciones estimadas para el conjunto NC se ordena el conjunto de mayor a menor con base en su calificación. Finalmente se eligen las 10 primeras películas y se muestran.

f. Describe tu sistema de recomendación como un agente inteligente haciendo explícita su REAS.

Agente	Rendimiento	Entorno	Actuadores	Sensores
Recomendador de películas.	Que al usuario le guste la recomendación.	Una base de datos.	Calificar películas basándose en la información de un usuario en particular.	Lector de base de datos.

g. ¿Cuál es el esquema que se utiliza para representar el conocimiento? y ¿Cuáles son las propiedades de ese esquema?

Debido al tipo de recomendación y al conjunto de datos recolectado, es suficiente y necesario una base de datos relacional ya que necesitamos que nuestro conocimiento sea fácil de actualizar y consultar, para ser más precisos con el sistema manejador MySQL.

Esta base de datos actúa como nuestra representación del conocimiento, donde el esquema de datos son los nombres de las tablas y sus atributos, además de los tipos de éstos (string, integer, etc). Este provee información sobre los datos (metadatos).

h. ¿Qué tipo esquema de representación de conocimiento es (declarativo o procedimental, etc.)? ¿Qué estrategias se pueden usar para manipular este tipo de conocimiento?

El esquema de representación de conocimiento es declarativo ya que se representa independientemente a su uso posterior, pero siendo más precisos este conocimiento es relacional ya que todo es expresable mediante tablas.

Para manipular este tipo de conocimiento necesitamos de procedimientos que lo enriquezcan ya que solo son datos y debemos transformarlos en información para nuestros propósitos. Para esto contamos con el lenguaje de consulta SQL, el cual es una estandarización donde tiene las siguiente ventajas:

- Tiene una sintaxis y semántica muy precisa.
- Lógica de primer orden más agregación.
- Cuenta con álgebra relacional.
- Restricciones de integridad.

i. Complejidad.

Complejidad espacial.

El algoritmo utiliza una matriz para hacer los cálculos, donde las filas representan las películas, las columnas representan a los usuarios y el valor de cada casilla representa un rating. Si el usuario no ha calificado una película simplemente se pone un 0 en la casilla. La complejidad de esta estructura es evidentemente $u \cdot p$, donde u es el número de usuarios y p es el número de películas. Por lo tanto la complejidad espacial es $O(u \cdot p)$.

Complejidad en tiempo.

Primero tomemos las siguientes definiciones:

- p : número de películas.

- u : número de usuarios.

- c : número de películas calificadas por el usuario x .

- n : número de películas no calificadas por el usuario x pero calificada por algún otro usuario.

Analicemos primero la complejidad de cada uno de los algoritmos. Para calcular la correlación de Pearson se necesitan hacer operaciones con vectores en R^n , las cuales son producto punto, promedio y magnitud. Cada vector es un arreglo asociado a las calificaciones de una película que le han dado los usuarios, entonces el número de entradas del vector es u . Como la complejidad de realizar cada una de estas operaciones es lineal y se utilizan un número constante de veces entonces la complejidad de Pearson es $O(u)$.

El algoritmo de estimar una calificación aplica Pearson por cada elemento calificado por el usuario x , es decir, c . Entonces la complejidad del algoritmo de estimar es $O(u * c)$.

El algoritmo de recomendación aplica la función de estimar por cada por cada película que no ha sido calificada por el usuario x , es decir, n , y después ordena usando quicksort basándose en el rating. Entonces la complejidad del algoritmo de recomendación es $O(u * c * n + n \log(n))$.

j. Ventajas y desventajas

Ventajas	Desventajas
<ul style="list-style-type: none">- Se apoya en otros usuarios para poder obtener la recomendación, haciendo esta inteligencia más humana.- Hace al programa más atractivo para los usuarios, ya que no siempre tienen que decidir desde 0 qué películas ver.	<ul style="list-style-type: none">- Los usuarios pueden no tener buen criterio para calificar películas.- Puede haber bots de empresas que se dediquen a darle buena calificación a sus películas.- Debido a la complejidad del algoritmo utilizado es difícil tratar con conjuntos de datos muy grandes.

4. Conclusiones

La forma actual en que almacenamos los datos nos permite tener un conocimiento o información mayor conforme va pasando el tiempo de las cosas que nos rodean y la relación que tienen con nosotros los usuarios. Gracias a esta sobreexplotación de datos se pueden usar diferentes métodos de filtrado para obtener nuevo conocimiento aunque para obtener esto es necesario iniciar con un conjunto no vacío de relaciones con los datos en caso de usar un filtrado colaborativo.

5. Referencias bibliográficas

RUSSELL, S. J.; NORVIG, P., (2004), *Inteligencia Artificial: Un enfoque moderno*, Segunda edición, Madrid, España, PEARSON EDUCACIÓN, S.A.

Su, X.; Khoshgoftaar, T. M., A Survey of Collaborative Filtering Techniques

Galán, S.. (-). Filtrado Colaborativo y Sistemas de Recomendación. 11 15/2017, de Universidad Carlos III de Madrid Sitio web: <http://www.it.uc3m.es/jvillena/irc/practicas/06-07/31.pdf>

<https://www.youtube.com/watch?v=h9gpufJFF-0>