

Fecha de publicación Mayo 8, 2020

Digital Object Identifier

# Sistema de Recomendación de películas

## AUTOR IVAN ARANGO<sup>1</sup>

<sup>1</sup>Facultad de ingeniería  
Universidad Rafael Landívar  
Email: ivanandres613@gmail.com

### ABSTRACT

Un sistema de recomendación es una herramienta que establece un conjunto de criterios y valoraciones sobre los datos de los usuarios para realizar predicciones sobre recomendaciones de elementos que puedan ser de utilidad o valor para el usuario. Estos sistemas seleccionan datos proporcionados por el usuario de forma directa o indirecta, y procede a analizar y procesar información del historial del usuario para transformar estos datos en conocimiento de recomendación. Los sistemas de recomendación en la actualidad tienen un nivel de eficiencia alto ya que pueden asociar elementos de nuestros perfiles de consumo como el historial de compras, selección de contenidos e inclusive nuestras horas de actividad, para realizar las recomendaciones. En el siguiente software a describir se utilizo el concepto de Content-Bases filtering, que consiste en aumentar las coincidencias entre valores de dos o más objetos mediante busquedas de propiedades similares. Se implemento dos estados del arranque frío, inicialmente por default el software tomó como criterio principal recomendar al usuario las películas con mejor puntuaciones otorgadas por IMDB, como segunda fase del arranque en frio se creo una biblioteca de películas en base a los criterios de búsqueda del usuario y proyectar recomendaciones más personalizadas en base a los gustos de los usuarios, alimentando una biblioteca de películas en base a las busquedas realizadas.

## I. INTRODUCCIÓN

En el presente artículo se analizaron y describieron las funcionalidades, recomendaciones y objetivos que cumplen los distintos tipos de algoritmos que se utilizan para el filtrado y recomendación de información en los sistemas de búsqueda. Los sistemas de recomendación sirven de ayuda para filtrar la información en base a los intereses de búsqueda del usuario y sugerir posibles intereses, además de reducir la sobrecarga de información en los sistemas al realizar dichas búsquedas, debido a que en la actualidad las plataformas manejan una alta cantidad de información por lo que puede causar retrasos en las búsquedas o colapsos de los servidores. Sin embargo, las áreas de machine learning permiten generar nuevas soluciones en base a los problemas descritos, como lo son los sistemas de recomendación. Como proyecto final del curso de Inteligencia artificial se generó un software simple de recomendación de películas en base a la clasificación de las 5000 mejores películas según IMDB, el cual consistía en tres fases, la de arranque frío por default, el arranque en frío en base a la data extraída de los usuarios y la búsqueda específica por el usuario. Este software fue implementado en el lenguaje de programación Java, utilizando la plataforma NetBeans 8.2 compatible con la versión Java SE Development

Kit 8u251 para Windows x64. Como propuesta final a la recomendación de películas, se creó un modelo probabilístico basado en el enfoque BCF que combina enfoques basados en elementos, en el cual la información obtenida del usuario se puede interpretar en un resultado específico. Mientras que el arranque en frío es un CF, que tiene como limitante que se necesita de una amplia data para poder recomendar de manera eficiente al usuario.

## II. CONCEPTOS

### A. SISTEMAS DE RECOMENDACIÓN

Un sistema de recomendación es un sistema inteligente que proporciona a los usuarios una serie de sugerencias personalizadas (recomendaciones) sobre un determinado tipo de elementos (ítems). Los sistemas de recomendación estudian las características de cada usuario y mediante un procesamiento de los datos, encuentran un subconjunto de ítems que pueden resultar de interés para el usuario. En los últimos años y debido principalmente a la sobre carga de información que tenemos en internet, han proliferado los sistemas de recomendación, los cuales proporcionan a los usuarios, información, productos, etc. que puedan ser de interés del usuario, tras realizar un "estudio" de su perfil,

sus gusto e incluso de la forma en la que el usuario navega por internet. Con estos datos es capaz de ofrecer al usuario productos o busquedas que le puedan interesar y que no sea el usuario el que se ponga a mirar los miles de productos que dispone la plataforma. Existen diferentes formas de realizar recomendaciones, en función de la navegación o compras de los usuarios o en función de lo que el usuario vote. A parte (y como es evidente) sabiendo las características de los usuarios (edad, sexo, profesión, etc.) también se podrían realizar recomendaciones en función de sus características como por ejemplo recomendar "tal producto" a hombres de entre 30 y 35 años o a mujeres de entre 18 a 25 años etc. Por tanto y con los ejemplos puesto, podemos pasar ya a clasificar los Sistemas de Recomendación, que se clasifican en 4 tipos:

- **SR con Filtrado basado en Contenido:** (ejm: YouTube) *Las recomendaciones se basan en el conocimiento que se tiene sobre los ítems que el usuario ha valorado (ya sea de forma implícita o explícita), y se le recomendarán ítems similares que le puedan gustar o interesar.*
- **SR con Filtrado Demográfico:** *Estas recomendaciones se realizan en función de las características de los usuarios (edad, sexo, situación geográfica, profesión, etc).*
- **SR basado en Filtrado Colaborativo:** (ejm: Filmaffinity) *El filtrado colaborativo consiste en ver que usuarios son similares al usuario activo (o usuario al que hay que realizarle las recomendaciones) y a continuación, recomendar aquellos ítems que no han sido votados por el usuario activo y que han resultado bien valorados por los usuarios similares.*
- **SR con métodos de Filtrado Híbrido:** (ejm: Amazon) *Mezclan alguno de los tres filtrados mencionados anteriormente para realizar recomendaciones e incluso lo combinan con alguna otra técnica de inteligencia artificial como pueda ser la lógica borrosa o la computación evolutiva.*

## B. CONCEPTO DE NORMALIZACIÓN

La normalización es una técnica que se aplica a menudo como parte de la preparación de datos para el aprendizaje automático. El objetivo de la normalización es cambiar los valores de las columnas numéricas del conjunto de datos para usar una escala común, sin distorsionar las diferencias en los intervalos de valores ni perder información. La normalización también es necesaria para que algunos algoritmos modelen los datos correctamente.

## III. TIPOS DE FILTRADO

El filtro basando en contenido, también conocido como, filtrado cognitivo recomienda elementos basados en una comparación entre el contenido de los elementos y un perfil de usuario. El contenido de cada elemento se representa

como un conjunto de descriptores o términos, generalmente las palabras que aparecen en un documento. El perfil del usuario se representa con los mismos términos y se construye analizando el contenido de los elementos que ha visto el usuario. Deben tenerse en cuenta varios problemas al implementar un sistema de filtrado basado en contenido. Primero, los términos pueden asignarse de forma automática o manual. Cuando los términos se asignan automáticamente, se debe elegir un método que pueda extraer estos términos de los elementos. En segundo lugar, los términos deben representarse de manera que tanto el perfil del usuario como los elementos puedan compararse de manera significativa. En tercer lugar, se debe elegir un algoritmo de aprendizaje que pueda aprender el perfil del usuario en función de los elementos vistos y pueda hacer recomendaciones basadas en este perfil de usuario.

### A. FILTRADO COLABORATIVO

El filtrado colaborativo se basa únicamente en las calificaciones de los usuarios u otro comportamiento del usuario, haciendo recomendaciones basadas en lo que les gustó o compró a los usuarios con un comportamiento similar. Una ventaja clave del filtrado colaborativo es que no depende del contenido del elemento y, por lo tanto, es capaz de recomendar elementos complejos como películas, sin comprender el elemento en sí.

### B. FILTRADO BASADO EN CONTENIDO

El filtrado basado en contenido, por otro lado, se basa en una descripción de elementos y un perfil de las preferencias del usuario combinadas de la siguiente manera. Primero, los ítems se describen con atributos, y para encontrar ítems similares, medimos la distancia entre los ítems utilizando una medida de distancia como la distancia cosenoide o el coeficiente de Pearson, el perfil del usuario ingresa a la ecuación. Dada la retroalimentación sobre el tipo de elementos que le gustan al usuario, podemos introducir pesos que especifiquen la importancia de un atributo de elemento específico.

### C. ENFOQUE HÍBRIDO

El filtrado colaborativo puede aprender las preferencias del usuario de las acciones del usuario con respecto a una fuente de contenido y usarlas en otros tipos de contenido. El filtrado basado en contenido se limita a recomendar contenido del mismo tipo que el usuario ya está utilizando.

### D. SIMILARIDAD COSENO

El algoritmo de similaridad del coseno es el encargado de desarrollar el cálculo de una medida de similitud entre dos vectores distintos de cero dentro de un espacio interno, donde se mide el coseno del ángulo entre ellos. Es vital comprender a profundidad el concepto de la similitud de coseno para entender su utilidad para nuestros proyectos. Esta se usa particularmente en un espacio positivo donde el resultado está claramente delimitado en [0,1][0,1].

### E. COEFICIENTE DE PEARSON

El coeficiente de correlación de Pearson es una prueba que mide la relación estadística entre dos variables continuas. Si la asociación entre los elementos no es lineal, entonces el coeficiente no se encuentra representado adecuadamente. El coeficiente de correlación puede tomar un rango de valores de +1 a -1. Un valor de 0 indica que no hay asociación entre las dos variables. Un valor mayor que 0 indica una asociación positiva. Es decir, a medida que aumenta el valor de una variable, también lo hace el valor de la otra. Un valor menor que 0 indica una asociación negativa; es decir, a medida que aumenta el valor de una variable, el valor de la otra disminuye.

### IV. DEFINICIÓN DE ALGORITMO

Como propuesta final del motor de recomendación de películas, se realizaron tres algoritmos para las distintas fases que presenta el software las cuales son:

- *Arranque en frío por default: Esta función se implemento cuando la plataforma aun no posee datos previamente solicitados por el usuario, por lo que se tomo el criterio de escoger las películas con mejor puntuación brindadas por IMDB, filtrando dichas películas y escogiendo de manera aleatoria 10 películas al usuario, para que pueda obtener diferentes recomendaciones al iniciar la aplicación.*
- *Recomendacion por medio del usuario: Se tomo como referencia el filtrado basado en contenido, asignando pesos de manera manual a los campos que se consideran que tiene mayor relevancia al describir una película (actor,director,año,pais,generos). Se almacenó la información brindada por el usuario en variables y se realizó una búsqueda para cada campo en la data de las películas, y al encontrar que la película en la posición i posee algún campo descrito por el usuario, se aplicaría el concepto de probabilidad condicional cuando se cumplen multiples eventos en un escenario descrito, y se calculó la probabilidad conjunta en base a los pesos asignados. Al tener la probabilidad de la posición i de cada película se almacenó en una lista que poseería las probabilidades calculadas previamente, para luego ordenarse de manera ascendente la lista y obtener los valores más bajos obtenidos, que equivaldrían a los valores que poseen mayor coincidencia a lo ingresado por el usuario. Al obtener estos dos valores se validó que películas poseían esta probabilidad y dependiendo del resultado de búsquedas obtenidas se mostraron en pantalla las recomendaciones personalizadas, si se obtuvieron más de 10 películas con esas coincidencias se tomaron 10 de manera aleatoria mientras que si se encontró menos de 10 se proyectaran todos los resultados encontrados. Al finalizar dicho proceso se almacenó dichas películas en un archivo de texto, que serviría como biblioteca de películas frequentadas por*

*los usuarios al realizar el arranque en frío.*

- *Recomendacion en frío en base a data del usuario: Al realizar esta recomendación se utiliza como referencia únicamente las películas escritas en el archivo de texto escrito en el proceso de recomendacion por medio del usuario, realizando nuevamente el algoritmo descrito en este proceso con la variante de ahorrar tiempo en no realizar búsquedas exhaustivas con mayor data, y realizar búsquedas más eficientes en data filtrada.*

### REFERENCES

- [1] Kaluza, Bostjan, *Machiner Learning in Java*. Birmingham, UK : Packt Publishing. 2016, chapter 6, pp.101.
- [2] *Algoritmo de similitud de coseno*,Graph Everywhere. [Online]. Available: <https://www.grapheverywhere.com/algoritmo-de-similitud-de-coseno/>
- [3] Kaluza, Bostjan, *Machiner Learning in Java*. Birmingham, UK : Packt Publishing. 2016, chapter 6, pp.103.