

## Sistema de Recomendación sobre películas

---

Javier Poyatos Amador  
Juan Luis Suárez Díaz  
jpoyatosamador@correo.ugr.es  
jlsuarezdiaz@correo.ugr.es  
76878633Y  
DNI Juanlu

29 de mayo de 2019

# Índice

<b>1</b>	<b>Introducción</b>	<b>3</b>
<b>2</b>	<b>Propuestas realizadas</b>	<b>3</b>
2.1	Recomendación basada en popularidad . . . . .	3
2.2	Recomendación basada en contenido . . . . .	4
2.3	Recomendación colaborativa . . . . .	4
2.4	Recomendación híbrida . . . . .	5
<b>3</b>	<b>Evaluación</b>	<b>5</b>

## 1. Introducción

En esta práctica se va a desarrollar un sistema de recomendación de películas en base a diferentes características. El conjunto de datos utilizados “The Movies Dataset” [1]. Éste proporciona bastante información acerca de las películas como puede ser los géneros que trata, presupuesto, título y lenguaje original, entre otros. El objetivo de esta práctica consiste en realizar un sistema de recomendación en base a diferentes criterios.

TODO : quizás una mini seccion explicando el tratamiento inicial del cjo de datos?

## 2. Propuestas realizadas

Las propuestas que se han llevado a cabo son las que se mencionan a continuación.

- **Recomendación basada en popularidad:** recomienda películas a un determinado usuario en función del número de votos y el voto medio junto con su puntuación de forma que estas películas superen un mínimo umbral de calidad.
- **Recomendación basada en contenido:** este sistema sugiere películas a un determinado usuario en función de aquellas de que el usuario haya visto y haya votado positivamente.
- **Recomendación colaborativa:** se recomienda películas a un usuario en función a otros usuarios que han visto películas similares a las que el primer usuario también ha visualizado.
- **Recomendación híbrida:** es una mezcla de los dos sistemas anteriores de forma que en base a un determinado usuario y a una (o varias) películas determina aquellas que pueden ser las que más le interesen a ese usuario.

### 2.1. Recomendación basada en popularidad

El objetivo de esta primera aproximación es simple: recomendar películas en base a una determinada métrica o sistema de puntuación. En este caso se muestran las  $n$  mejores películas en función del voto medio, votos totales y de la puntuación de cada película.

Para realizar tal tarea simplemente necesitamos saber los votos de cada película, el número de votos medio y tras ello mostramos aquellas películas que superen un determinado cuantil, que representa el mínimo número de votos que necesita una película para estar dentro de la lista de recomendaciones. Con estos dos coeficientes calculamos un peso medio para cada película que va a ser su puntuación

$$WR = \frac{vR}{v + m} + \frac{mC}{v + m}$$

donde  $v$  es el número de votos para cada película,  $m$  es el mínimo número de votos que necesita una película para estar en las recomendaciones,  $R$  es el rating medio de la película y  $C$  la media de esa película en función del informe de la película.

Un ejemplo de este primer sistema de recomendación se muestra en la Figura ??.

## 2.2. Recomendación basada en contenido

Uno de los problemas que tiene la aproximación anterior es que distintos usuarios van a obtener la misma recomendación cuando no tiene por qué darse que a ambos les gusten las mismas películas: a uno le pueden gustar las románticas y a otro las comedias.

El objetivo de nuestro sistema basado en contenido es personalizar la recomendación en base a las películas que el usuario haya visto y valorado positivamente. Para ello, el sistema calcula la similitud entre las películas en base a una cierta métrica y sugiere, para un determinado usuario, un número de películas acorde a sus gustos.

Dependiendo de la métrica usada se obtendrán resultados distintos. En nuestro caso hemos fijado como métrica de similitud la métrica del coseno. Ahora bien, previo a aplicar esta métrica proponemos dos métodos de similitud de las películas: método basado en las descripciones de las películas y método basado en créditos, géneros y palabras clave.

- Método de similitud basado en descripciones de las películas: creamos un objeto TF-IDF según las palabras y en base a éste calculamos la similitud de las películas con el coseno.
- Método basado en créditos, géneros y palabras clave: con estos tres datos de nuestro conjunto de datos, creamos la matriz para calcular la similitud.

Una vez que hemos calculado esta métrica simplemente hemos de tomar las mejores películas según el valor para cada una de ella según la métrica para terminar mostrando las  $n$  mejores. El resultado se puede ver en la Figura ??.

## 2.3. Recomendación colaborativa

El problema que tiene el sistema anterior no puede recomendar películas de diferentes géneros ya que solamente puede recomendar películas del mismo tipo, es decir, que sean cercanas en función de la métrica. El objetivo de esta nueva aproximación es que si dos usuarios A y B han visto las películas 1 y 2 y ambos las valoran positivamente y A ve la película 3, entonces a B le puede interesar también 3.

Para ello hemos usado dos métodos para minimizar la métrica RMSE y son SVD y KNN [2]. Lo primero que hace el sistema es tomar las películas que el usuario no ha visto y sobre ellas predecir las que les pueden interesar más. Para ello usamos el rating del restos de usuarios y el rating del propio usuario y vemos cuáles son los identificadores de

películas no vistas por éste. Tras ello, localizamos el título, identificador y total de votos y votos medio de cada una de ellas para acabar calculando una estimación en función al modelo aprendido. El resultado de esta nueva aproximación se muestra en la Figura ??.

## **2.4. Recomendación híbrida**

Esta aproximación es una mezcla de los dos anteriores. En este caso nuestro sistema va a procesar un usuario y una lista de películas. El proceso para la aproximación híbrida consiste en calcular una métrica de similitud como se expuso en el método basado en contenido y, tras ello, obtener las mejores películas que nos da esa aproximación (las  $n$  mejores). A continuación, para esas películas usamos el modelo aprendido por KNN o SVD y obtenemos las predicciones.

## **3. Evaluación**

### **Referencias**

[1] <https://www.kaggle.com/rounakbanik/the-movies-dataset>

[2] <http://surpriselib.com/>