

Taller de sistemas de recomendación por filtro colaborativo

Objetivo

El propósito de este taller es crear un modelo de recomendación basado en filtro colaborativo paso por paso en Excel, para poder entender cada etapa del proceso. Trataremos con una base de 25 usuarios que han calificado 100 películas (ítems), con ratings de 1 a 5. Ningún usuario ha calificado las 100 películas, y ninguna película ha sido calificada por los 25 usuarios, por lo que hay ratings a predecir.

Parte 1: Entendimiento de los datos

Todos los datos están en las dos primeras pestañas; en la primera están con las películas en las filas, en la segunda están los usuarios en las filas. En un recomendador no personalizado, ¿qué películas recomendaría usted?

¿Cuáles son las películas menos queridas?

¿Cuáles son las películas más polémicas (con mayor dispersión)?

¿Cuáles son los usuarios que notan mejor las películas? ¿los que peor?

¿Cuáles son los usuarios con mayor dispersión de ratings?

Parte 2: Creación de matriz de similitud

El hecho de repetir los datos transpuestos de esta manera hace que sea más fácil el cálculo de la correlación.

En una nueva hoja de Excel llamada “Correl”, crear una matriz de similitud entre usuarios, utilizando la correlación de Pearson como medida de similitud, con la función “COEF.DE.CORREL”.

Parte 3: Afinidad de las películas, sin normalizar datos

Vamos a centrarnos en el usuario 3867. En la pestaña “Scores”, en la columna A, ordenar los usuarios desde el más cercano hasta el menos cercano al 3867, teniendo en cuenta la correlación entre usuarios calculada en la pestaña “Correl”. Completar la tabla con los ratings dados originalmente por los usuarios, utilizando las funciones “BUSCARV” y “ESBLANCO” (No queremos que películas sin ratings presenten un rating de 0).

En las filas 29, 30 y 31 calcular el score de cada película para el usuario 3867 sin tener en cuenta la similitud de los usuarios:

- Fila 29: los 5 usuarios más cercanos, hayan o no hayan notado la película. Calcular el promedio de los ratings. Las películas no notadas no serán tenidas en cuenta.
- Fila 30: los usuarios que superen el umbral de correlación de 0.3. Calcular el promedio de los ratings. Las películas no notadas no serán tenidas en cuenta.
- Fila 31: los 5 usuarios más cercanos que hayan notado la película, y cuya correlación es positiva. Calcular el promedio de los ratings. Si no hay suficientes usuarios cercanos, no se podrá estimar el rating del usuario 3867.

En las filas 32 y 33 calcular el score de cada película para el usuario 3867 teniendo en cuenta la similitud de los usuarios, teniendo en cuenta respectivamente:

- Fila 32: los 5 usuarios más cercanos, hayan o no hayan notado la película. Calcular el promedio de los ratings. Las películas no notadas no serán tenidas en cuenta.
- Fila 33: los usuarios que superen el umbral de correlación de 0.3. Calcular el promedio de los ratings. Las películas no notadas no serán tenidas en cuenta.

Escoger las 5 películas a recomendar para cada una de las técnicas.

Parte 4: Afinidad de las películas, datos normalizados

En la pestaña “Scores”, vamos a completar la columna CY con los promedios de notas dados por cada usuario.

En la pestaña “ScoresNorm”, vamos a sustraer al valor de rating dado por cada usuario su promedio de notas general.

En las filas 29, 30 y 31 calcular el score de cada película para el usuario 3867 sin tener en cuenta la similitud de los usuarios:

- Fila 29: los 5 usuarios más cercanos, hayan o no hayan notado la película. Calcular el promedio de los ratings centrados y agregarle el promedio de las notas del usuario 3867. Las películas no notadas no serán tenidas en cuenta.
- Fila 30: los usuarios que superen el umbral de correlación de 0.3. Calcular el promedio de los ratings centrados y agregarle el promedio de las notas del usuario 3867. Las películas no notadas no serán tenidas en cuenta.
- Fila 31: los 5 usuarios más cercanos que hayan notado la película, y cuya correlación es positiva. Calcular el promedio de los ratings centrados y agregarle el promedio de las notas del usuario 3867. Si no hay suficientes usuarios cercanos, no se podrá estimar el rating del usuario 3867.

En las filas 32 y 33 calcular el score de cada película para el usuario 3867 teniendo en cuenta la similitud de los usuarios, teniendo en cuenta respectivamente:

- Fila 32: los 5 usuarios más cercanos, hayan o no hayan notado la película. Calcular el promedio de los ratings centrados y agregarle el promedio de las notas del usuario 3867. Las películas no notadas no serán tenidas en cuenta.
- Fila 33: los usuarios que superen el umbral de correlación de 0.3. Calcular el promedio de los ratings centrados y agregarle el promedio de las notas del usuario 3867. Las películas no notadas no serán tenidas en cuenta.

Escoger las 5 películas a recomendar para cada una de las técnicas en la pestaña “Preds”