

A continuación, explicamos las técnicas usadas










## ¿Cómo funciona Collaborative Filtering?

Para explicar cómo funciona Collaborative Filtering vamos a entender cómo será el dataset.

### Ejemplo de Dataset

Necesitaremos, «ítems» y las valoraciones de los usuarios. Los ítems pueden ser, canciones, películas, productos, ó lo que sea que queremos recomendar.

Entonces nos quedará una matriz de este tipo, donde la intersección entre fila y columna es una valoración del usuario:

					
	2	5		4	
		1	5		3
	5	3	4	3	
		3	5	1	1

En esta «gráfica educativa» tenemos una matriz con productos (a la izquierda) y los ítems (arriba). En este ejemplo los ítems serán frutas y cada celda contiene la valoración hecha por cada usuario de ese ítem. Las casillas vacías significa que el usuario aún no ha probado esa fruta.

Entonces veremos que tenemos «huecos» en la tabla pues evidentemente no todos los usuarios tienen o «valoraron» todos los ítems. Por ejemplo si los ítems fueran «películas», es evidente que un usuario no habrá visto <<todas las películas del mundo>>... entonces

esos huecos son justamente los que con nuestro algoritmo «rellenaremos» para recomendar ítems al usuario.

Una matriz con muchas celdas vacías se dice -en inglés- que es **sparse** (y suele ser normal) en cambio si tuviéramos la mayoría de las celdas cubiertas con valoraciones, se llamará **dense**.

## Tipos de Collaborative Filtering

- **User-based:** (Este es el que veremos a continuación)
  - Se identifican usuarios similares
  - Se recomiendan nuevos ítems a otros usuarios basado en el rating dado por otros usuarios similares (que no haya valorado este usuario)
- **Item-based:**
  - Calcular la similitud entre ítems
  - Encontrar los «mejores ítems similares» a los que un usuario no tenga evaluados y recomendárselos.

## Predecir gustos (User-based)

Collaborative Filtering intentará encontrar usuarios similares, para ofrecerle ítems «bien valorados» para ese perfil en concreto (lo que antes llamé «rellenar los huecos» en la matriz). Hay diversas maneras de medir ó calcular la similitud entre usuarios y de ello dependerá que se den buenas recomendaciones. Pero tengamos en cuenta que estamos hablando de buscar similitud entre «gustos» del usuario sobre esos ítems, me refiero a que no buscaremos perfiles similares por ser del mismo sexo, edad ó nivel educativo. Sólo nos valdremos de los ítems que ha experimentado, valorado (y podría ser su secuencia temporal) para agrupar usuarios «parecidos».

Una de las maneras de medir esa similitud se llama «**distancia por coseno de los vectores**» y por simplificar el concepto, digamos que crea un espacio vectorial con n dimensiones correspondientes a los n ítems y sitúa los vectores siendo su medida el «valor rating» de cada usuario -a ese ítem-. Luego calcula el ángulo entre los vectores partiendo de la «coordenada cero». A «poca distancia» entre ángulos, se corresponde con usuarios con mayor similitud.

Este método no es siempre perfecto... pero es bastante útil y rápido de calcular.

## Calcular los Ratings

Una vez que tenemos la matriz de similitud, nos valdremos de otra operación matemática para calcular las recomendaciones.

$$R_U = \left( \sum_{u=1}^n R_u * S_u \right)$$

FORMULA para calcular los ratings faltantes: sería algo así como «Matriz de similitud PROD. VECTORIAL ratings / (sumatoria de cada fila de ratings) Transpuesta  
Lo haremos es: cada rating se multiplica por el factor de similitud de usuario que dio el rating. La predicción final por usuario será igual a la suma del peso de los ratings dividido por la «suma ponderada».