

THE BRIDGE

*Sistemas de
Recomendación*

¿Qué es un sistema de recomendación?

- Soluciones que realizan recomendaciones para emparejar usuarios con productos
- El objetivo es ayudar a los usuarios a tomar decisiones frente a grandes cantidades de información, mejorando la experiencia de usuario y reduciendo el tiempo empleado en la selección de productos a consumir.

amazon
ebay YAHOO!

You Tube
IMDb NETFLIX

The New York Times



lost.fm Spotify

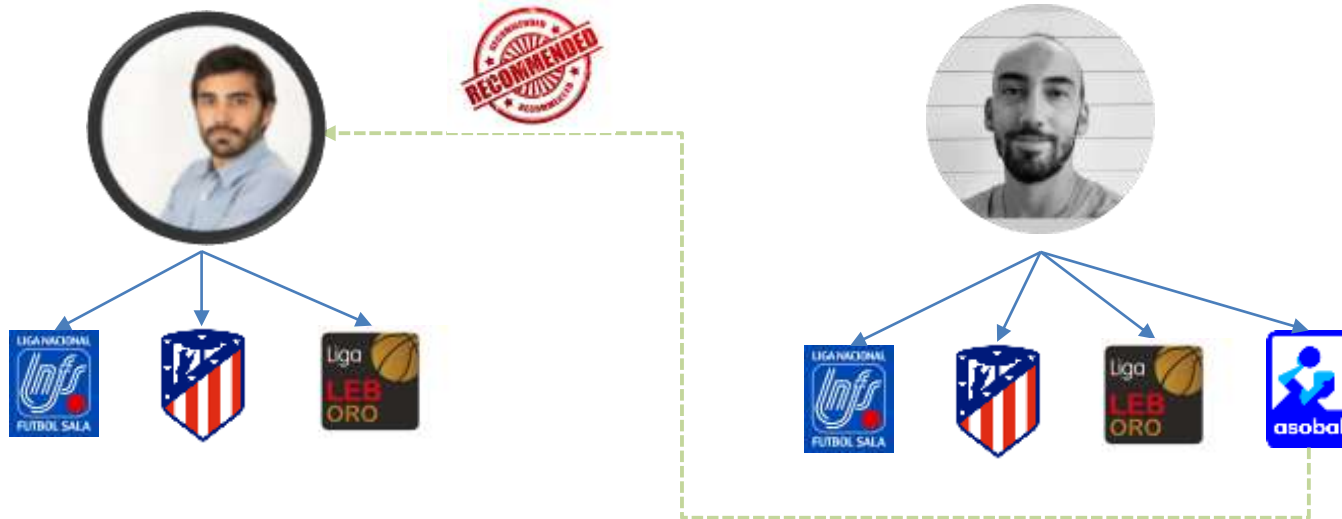


Filtrado colaborativo

- Utilizan a usuarios con gustos afines u objetos similares para estimar recomendaciones.

User – User

Estos algoritmos realizan recomendaciones utilizando los consumos de otros usuarios y la afinidad entre éstos.



Filtrado colaborativo

- Utilizan a usuarios con gustos afines u objetos similares para estimar recomendaciones.

Item-Item

Estos algoritmos realizan recomendaciones utilizando los consumos del usuario y la afinidad entre los items



Filtrado colaborativo

¿Cómo medimos la similitud entre items?



Película	Acción	Comedia
La Máscara	6	6
La Vida de Brian	1	10
Misión Imposible	10	2

Filtrado colaborativo

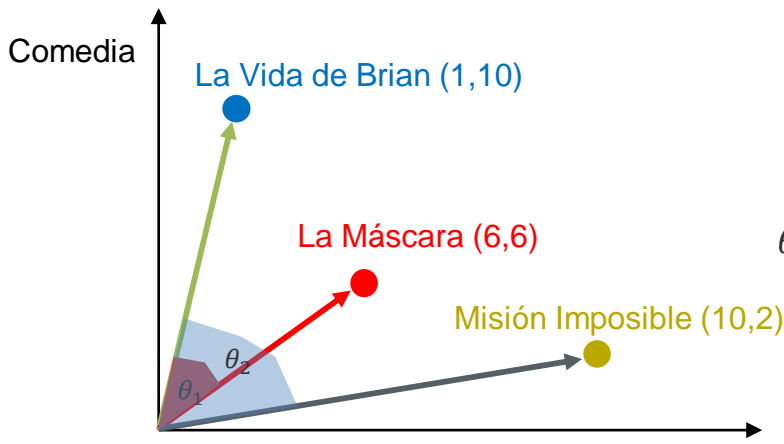
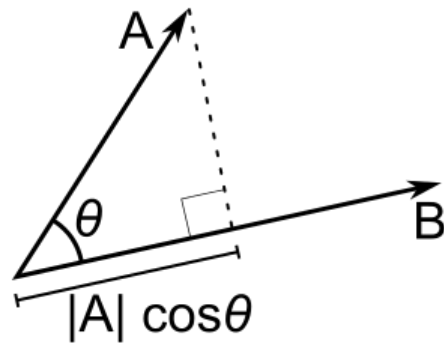
• ¿Cómo medimos la similitud entre items?

Similitud del coseno

$$\text{similitud}(A, B) = \cos(\theta) = \frac{v_A v_B}{\|v_A\| \|v_B\|} \in [0, 1]$$

A y B son los objetos

v_A y v_B son los vectores que representan los objetos A y B



$\theta_1 < \theta_2 \Rightarrow \cos(\theta_1) > \cos(\theta_2)$
 \Rightarrow La Vida de Brian es más similar a La Máscara que a Misión Imposible

Filtrado colaborativo

¿Cómo medimos la similitud entre items?

Película	Acción	Comedia	v	$\ v\ = \sqrt{x^2 + y^2}$
La Máscara	6	6	(6,6)	8.48
La Vida de Brian	1	10	(1,10)	10.05
Misión Imposible	10	2	(10,2)	10.2

$$\text{similitud}(\text{"La Máscara"}, \text{"La Vida de Brian"}) = \frac{v_A v_B}{\|v_A\| \|v_B\|} = \frac{6 \cdot 1 + 6 \cdot 10}{8.48 \cdot 10.05} = 0.7$$

$$\text{similitud}(\text{"La Vida de Brian"}, \text{"Misión Imposible"}) = \frac{v_A v_B}{\|v_A\| \|v_B\|} = \frac{1 \cdot 10 + 10 \cdot 2}{10.05 \cdot 10.2} = 0.29$$

Recomiendo ítems similares a los que te gustan

Filtrado colaborativo



VideoID	Rating
A	4
B	1
C	3
D	???

¿Cuál es el rating esperado del vídeo D?

Matriz de similitudes

	A	B	C	D
A	1	0	0.4	0.6
B	0	1	0	0
C	0.4	0	1	0.9
D	0.6	0	0.9	1

$$\begin{aligned} \text{Rating}(\text{Rafa}, D) &= \text{rating}_A \times \text{similitud}(A, D) \\ &\quad + \text{rating}_B \times \text{similitud}(B, D) \\ &\quad + \text{rating}_C \times \text{similitud}(C, D) \end{aligned}$$

$$\begin{aligned} &= 4 \times 0.6 \\ &\quad + 1 \times 0 \\ &\quad + 3 \times 0.9 = \mathbf{5.1} \end{aligned}$$

Haciendo este cálculo para cada vídeo no visualizado, podemos establecer un ranking de los vídeos con mayor estimación de rating

Filtrado colaborativo

¿Cómo medimos la similitud entre usuarios?

Coeficiente de correlación $\in [-1,1]$

	La Máscara	Titanic	El Padrino	Misión Imposible	The Disaster Artist	Media
Rafa	2	3	5	3	?	3.25
Antonio	5	5	1	1	4	3.2
Raúl	2	4	4	4	5	3.8
Guillermo	2	5	3	4	3	3.4

Recomendable?

$\text{similitud}(\text{"Rafa"}, \text{"Antonio"}) = -0.68$

$\text{similitud}(\text{"Rafa"}, \text{"Raúl"}) = 0.66$

$\text{similitud}(\text{"Rafa"}, \text{"Guillermo"}) = 0.1$

Recomiendo ítems que le gustan a usuarios similares a ti (similitud>0), ponderando con el grado de similitud

$$\begin{aligned}\text{predicción}(\text{"Rafa"}, \text{"The Disaster Artist"}) &= 3.25 + \frac{0.66 \cdot (5 - 3.8) + 0.1 \cdot (3 - 3.4)}{0.66 + 0.1} = \\ &= 3.25 + 0.99 = 4.2\end{aligned}$$



Filtrado colaborativo

La factorización consiste en expresar un problema grande como producto de factores más pequeños, e.g.

$$55 = 5 \times 11$$

Factorización matricial

	La Máscara	Titanic	El Padrino	Misión Imposible	The Disaster Artist
Rafa	2	3	?	3	?
Antonio	?	5	1	1	4
Raúl	2	4	4	4	?
Guillermo	2	?	3	?	3

$$\begin{pmatrix} 2 & 3 & ? & 3 & ? \\ ? & 5 & 1 & 1 & 4 \\ 2 & 4 & 4 & 4 & ? \\ 2 & ? & 3 & ? & 3 \end{pmatrix} \approx \begin{pmatrix} U \end{pmatrix} \begin{pmatrix} S \end{pmatrix} \begin{pmatrix} V \end{pmatrix}$$

$$\begin{pmatrix} 0.40 \\ 0.55 \\ 0.68 \\ 0.27 \end{pmatrix} \cdot (9.88) \cdot (0.27 \quad 0.68 \quad 0.41 \quad 0.45 \quad 0.31) = \begin{pmatrix} 1.1 & 2.7 & 1.6 & 1.8 & 1.2 \\ 1.5 & 3.7 & 2.3 & 2.5 & 1.7 \\ 1.8 & 4.5 & 2.8 & 3 & 2 \\ 0.7 & 1.8 & 1.1 & 1.2 & 0.8 \end{pmatrix}$$

Recomiendo ítems con mayor valor en la matriz

Problemas frecuentes

- *Cold-start*: Un usuario/item nuevo no tiene suficiente información como para asignarle recomendaciones
- *Sparsity*: Muy poca gente califica los ítems
- *Transparency*: No se pueden explicar los motivos de las recomendaciones

Técnica	Ventajas	Desventajas
User-User	<ul style="list-style-type: none">• Fácil de implementar• Diversidad en las recomendaciones	<ul style="list-style-type: none">• Sparsity• Lento• Cold start para usuarios e ítems
Item-Item	<ul style="list-style-type: none">• La recomendación sólo depende de los gustos del usuario• Puedo recomendar ítems que no han sido puntuados por usuarios• Menos computación (ítems << users)	<ul style="list-style-type: none">• Sparsity• Menor diversidad en las recomendaciones• Necesidad de etiquetado
Factorización	<ul style="list-style-type: none">• Más rápido	<ul style="list-style-type: none">• Transparency



¡Gracias!