Sistemas de Recomendación

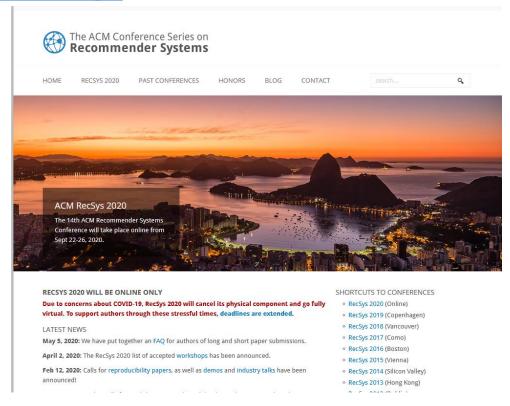
Referencias

• Introduction to Recommender SystemsHandbook Ricci, F., Rokach, L., & Shapira, B. (2011). Springer, Boston, MA

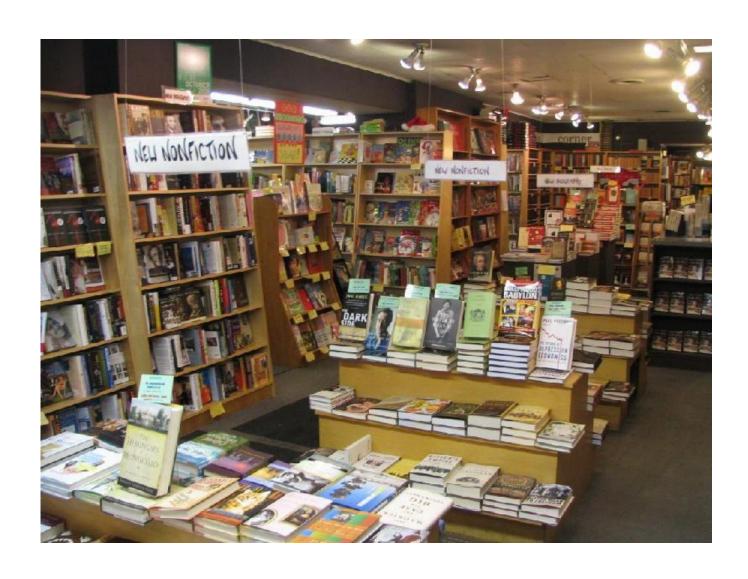
• A Programmer's Guide to Data Mining. Zacharski, R. (2017).

ACM Recommender System

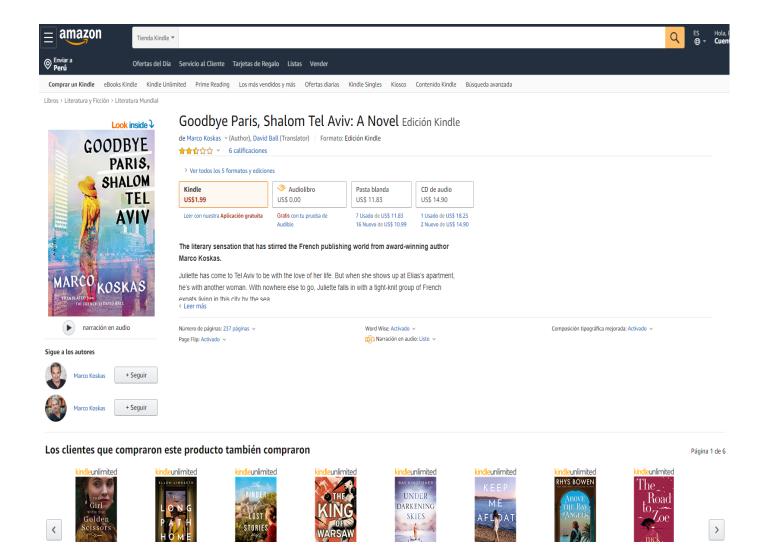
https://recsys.acm.org/



Recomendación Tradicional



Sistemas de Recomendación Actuales



Me gusta lo que te gusta?



Learn How You Can Improve User Experience by Eliminating Unnecessary Learning and Confusion for Your Users



262 comentarios 980 veces compartido



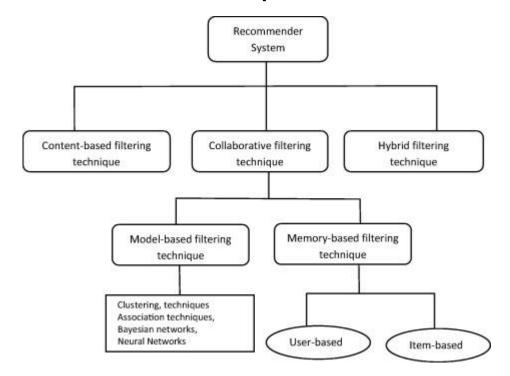
GamesRadar+ ②

...

...

Sistema de Recomendación

 Sugiere productos, servicios e información a los usuarios en función al análisis de datos generados dinámicamente de acuerdo con las preferencias, intereses o comportamiento.



Isinkaye, F. O., Folajimi, Y. O., & Ojokoh, B. A. (2015). Recommendation systems: Principles, methods and evaluation. *Egyptian Informatics Journal*, *16*(3), 261-273.

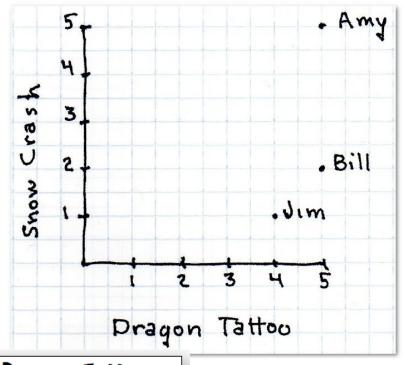
Filtro Colaborativo

- Realizar recomendaciones basadas en otras personas.
- Las personas colaboran para formular recomendaciones
 - Ejemplo: Comprar un libro, comprar un juego, etc
 - Buscar a una persona que tenga gustos similares a los tuyos
 - Una vez que se encuentra a esa persona, puedo ver lo que le gusta
 - Se recomienda el producto.

Cómo encuentro que alguien es similar?

Recomendar un libro al usuario X

	Snow Crash	Girl with the Dragon Tattoo
Amy	5☆	5☆
Bill	2☆	5☆
Jim	1☆	4 ☆



	Snow Crash	Girl with the Dragon Tattoo
Amy	5☆	5☆
Bill	2☆	5☆
Jim	1☆	4☆
Х	4	2

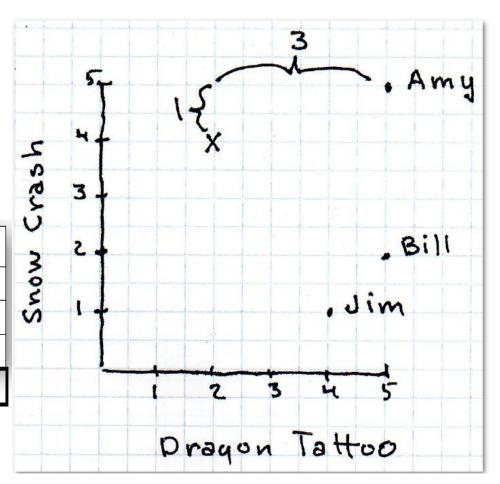
Calcular la Distancia

Distancia de Manhattan

$$| x_1 - x_2 | + | y_1 - y_2 |$$

La distancia entre Amy y Ux

	Snow Crash	Girl with the Dragon Tattoo		
Amy	5☆	5☆		
Bill	2☆	5☆		
Jim	1☆	4 ☆		
Χ	4	2		



Distancia de Manhattan

Dm(A,x) =
$$|5-4| + |5-2|$$

1 + 3
4

	Distance from Ms. X
Amy	4
Bill	5
Jim	5

Distancia Euclidiana

$$\sqrt{(x_1-x_2)^2+(y_1-y_2)^2}$$

	Snow Crash	Girl with the Dragon Tattoo
Amy	5☆	5☆
Bill	2☆	5☆
Jim	1☆	4 ☆
X	4	2

$$\sqrt{(5-2)^2 + (5-4)^2} = \sqrt{3^2 + 1^2} = \sqrt{10} = 3.16$$

	Distance from Ms. X	
Amy	3.16	
Bill	3.61	
Jim	3.61	

N-Dimensional

	Angelica	Bill	Chan	Dan	Hailey	Jordyn	Sam	Veronica
Blues Traveler	3.5	2	5	3	-	-	5	3
Broken Bells	2	3.5	1	4	4	4.5	2	-
Deadmau5	-	4	1	4.5	1	4	-	-
Norah Jones	4.5	-	3	-	4	5	3	5
Phoenix	5	2	5	3	-	5	5	4
Slightly Stoopid	1.5	3.5	1	4.5	-	4.5	4	2.5
The Strokes	2.5	-	-	4	4	4	5	3
Vampire Weekend	2	3	-	2	1	4	-	-

Calcular la Distancia de Manhattan entre Angelica y Bill

	Angelica	Bill	Difference
Blues Traveler	3.5	2	1.5
Broken Bells	2	3.5	1.5
Deadmau5	-	4	
Norah Jones	4.5	-	
Phoenix	5	2	3
Slightly Stoopid	1.5	3.5	2
The Strokes	2.5	-	-
Vampire Weekend	2	3	1
Manhattan Distance:			9

Calcular la Distancia Euclidiana entre Angelica y Bill

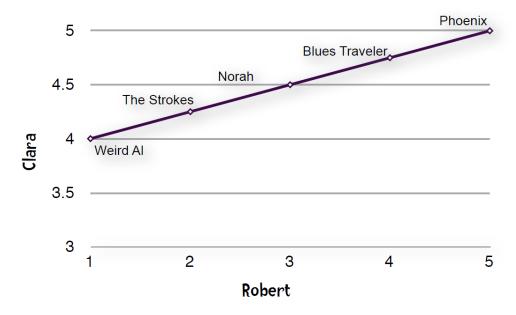
	Angelica	Bill	Difference	Difference ²
Blues Traveler	3.5	2	1.5	2.25
Broken Bells	2	3.5	1.5	2.25
Deadmau5	-	4		
Norah Jones	4.5	-		
Phoenix	5	2	3	9
Slightly Stoopid	1.5	3.5	2	4
The Strokes	2.5	-	-	
Vampire Weekend	2	3	1	1
Sum of squares				18.5
Euclidean Distance				4.3

$$=\sqrt{18.5}=4.3$$

Coeficiente de Correlación de Pearson

	Blues Traveler	Norah Jones	Phoenix	The Strokes	Weird Al
Clara	4.75	4.5	5	4.25	4
Robert	4	3	5	2	1

Usuarios tienen diferentes escalas



Coeficiente de Correlación de Pearson

- Es una medida de correlación entre dos variables
- Varia entre -1 y 1
 - 1 acuerdo perfecto
 - -1 desacuerdo perfecto

$$r = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \overline{x})^2} \sqrt{\sum_{i=1}^{n} (y_i - \overline{y})^2}}$$

Coeficiente de Correlación de Pearson

• Aproximación de Pearson:

$$r = \frac{\sum_{i=1}^{n} x_{i} y_{i} - \frac{\sum_{i=1}^{n} x_{i} \sum_{i=1}^{n} y_{i}}{n}}{\sqrt{\sum_{i=1}^{n} x_{i}^{2} - \frac{(\sum_{i=1}^{n} x_{i})^{2}}{n}} \sqrt{\sum_{i=1}^{n} y_{i}^{2} - \frac{(\sum_{i=1}^{n} y_{i})^{2}}{n}}}$$

Aproximación de Pearson

$$r = \frac{\sum_{i=1}^{n} x_{i} y_{i} - \frac{\sum_{i=1}^{n} x_{i} \sum_{i=1}^{n} y_{i}}{n}}{\sqrt{\sum_{i=1}^{n} x_{i}^{2} - \frac{(\sum_{i=1}^{n} x_{i})^{2}}{n}} \sqrt{\sum_{i=1}^{n} y_{i}^{2} - \frac{(\sum_{i=1}^{n} y_{i})^{2}}{n}}}$$

	Blues Traveler	Norah Jones	Phoenix	The Strokes	Weird Al
Clara	4.75	4.5	5	4.25	4
Robert	4	3	5	2	1

Robert 4 3 5 2 1
$$\sum_{i=1}^{n} x_{i} y_{i} = (4.75 \times 4) + (4.5 \times 3) + (5 \times 5) + (4.25 \times 2) + (4 \times 1) = 70$$

$$\sum_{i=1}^{n} x_{i} \sum_{i=1}^{n} y_{i} = 4.75 + 4.5 + 5 + 4.25 + 4 = 22.5$$

$$\sum_{i=1}^{n} x_{i} \sum_{i=1}^{n} y_{i} = 4 + 3 + 5 + 2 + 1 = 15$$

$$= \frac{22.5 \times 15}{5} = 67.5$$

$$\sum_{i=1}^{n} x_{i}^{2} = (4.75)^{2} + (4.5)^{2} + (5)^{2} + (4.25)^{2} + (4)^{2} = 101.875$$

$$\sum_{i=1}^{n} x_{i}^{2} = \frac{22.5^{2}}{5}$$

$$\sum_{i=1}^{n} x_{i}^{2} = \frac{22.5^{2}}{5}$$

$$\sqrt{\sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}} = \sqrt{101.875 - 101.25} = \sqrt{.625} = .79057$$

$$\sqrt{\sum_{i=1}^{n} y_i^2 - \frac{(\sum_{i=1}^{n} y_i)^2}{n}} = \sqrt{55 - 45} = 3.162277$$

$$r = \frac{2.5}{.79057(3.162277)} = \frac{2.5}{2.5} = 1.00$$

Similitud del Coseno

$$\cos(x,y) = \frac{x \cdot y}{\|x\| \times \|y\|}$$
. Producto Punto
$$\|x\| = \text{Longitud del Vector}$$

	Blues Traveler	Norah Jones	Phoenix	The Strokes	Weird Al
Clara	4.75	4.5	5	4.25	4
Robert	4	3	5	2	1

$$x \cdot y = (4.75 \times 4) + (4.5 \times 3) + (5 \times 5) + (4.25 \times 2) + (4 \times 1) = 70$$

$$||x|| = \sqrt{4.75^2 + 4.5^2 + 5^2 + 4.25^2 + 4^2} = \sqrt{101.875} = 10.09$$

$$||y|| = \sqrt{4^2 + 3^2 + 5^2 + 2^2 + 1^2} = \sqrt{55} = 7.416$$

$$\cos(x,y) = \frac{70}{10.093 \times 7.416} = \frac{70}{74.85} = 0.935$$

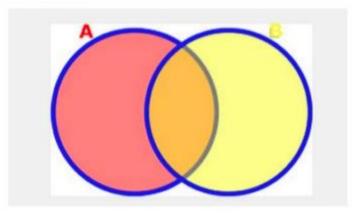
1= similitud perfecta

-1 = negativa similitud perfecta

Distancia Jaccard

• El índice de Jaccard (coeficiente de similitud de Jaccard) mide la similitud entre dos conjuntos

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|}.$$



Distancia Jaccard

- El índice de Jaccard (coeficiente de similitud de Jaccard) mide la similitud entre dos conjuntos
- La medida de distancia de Jaccard mide la disimilitud y es definida como:

$$J_{\delta}(A,B) = 1 - J(A,B) = \frac{|A \cup B| - |A \cap B|}{|A \cup B|}$$

Distancia Jaccard

 El índice de Jaccard (coeficiente de similitud de Jaccard) mide la similitud entre dos conjuntos

	HP_1	HP_2	HP_3	TW1	SW1	SW2	<i>SW</i> 3
A	4			5	1		
В	5	5	4				
C				2	4	5	
D		3					3

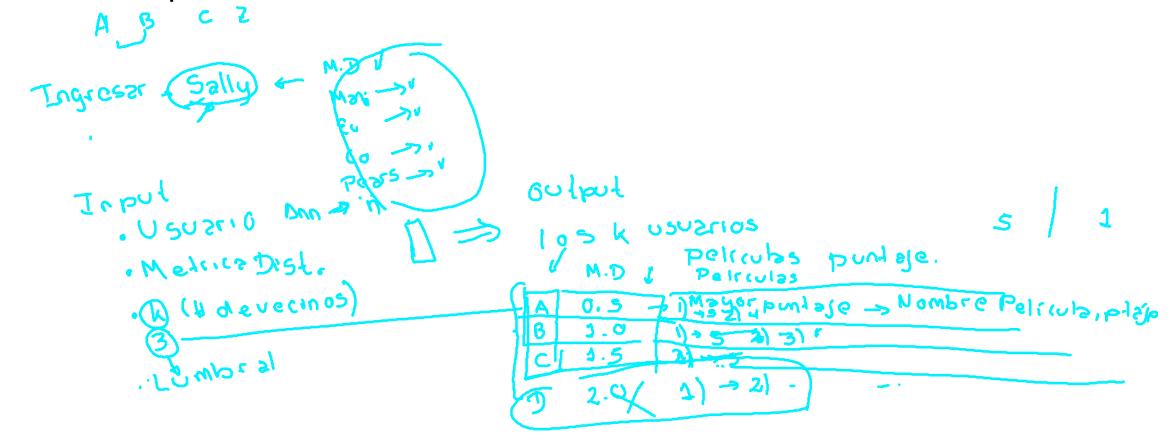
- J(A, B) = $\frac{A \cap B}{A \cup B} = \frac{1}{5}$ La similitud es 1/5, y la distancia es 4/5
- J(A, C) = $\frac{A \cap C}{A \cup C} = \frac{2}{4}$ La similitud es 2/4, y la distancia es 1/2
- N, U

Qué medida de similitud usar?

- Manhattan, Euclidiana
 - Si los datos son densos (valores no ceros) y la magnitud del valor del atributo es importante.
- Coseno
 - Si los datos son esparzos
- Pearson
 - Cuando los usuarios tienen diferentes escalas

K-nearest neighbor

• En el contexto de filtro colaborativo, se usa las k personas más cercanas para determinar más cercanos



K-nn: Ejemplo

- Hacer una recomendación para Ann usando el k-nn para k=3
 - Cuál es el grado de influencia que tiene cada persona?

Person Person	Pearson	Puntaje	
Sally	8.0	3.5	
Eric	0.7	5.0	
	1.5		

El puntaje proyectado para Sally y Erick =

$$3.5 * (0.8/1.5) + 5 (0.7/1.5) = 4.2$$

K-nn: Ejemplo

Suponga que Amanda, Erick y Sally, dan el puntaje a la banda Grey

Per <i>s</i> on	Grey Wardens Rating
Amanda	4.5
Eric	5
Sally	3.5

Cuál es el puntaje proyectado para la banda Grey para los 3-knn

Person	Grey Wardens Rating	Influencia
Amanda	4.5	25%
Eric	5	35%
Sally	3.5	40%

% Proyectado = 4.5*0.25 + 5*0.35 + 3.5 * 0.4 = 4.275

K-nn: Ejemplo

• Si k=2, cuál es el puntaje para la banda Grey

Per <i>s</i> on	Grey Wardens Rating
Amanda	4.5
Eric	5
Sally	3.5

• Cuál es el puntaje proyectado para la banda Grey para los 3-knn

Person	Grey Wardens Rating	Influencia
Amanda	4.5	25%
Eric	5	35%
Sally	3.5	40%

[%] Proyectado = 4.5*0.25 + 5*0.35 + 3.5 * 0.4 = 4.275