UNIVERSIDAD
NACIONAL
DE COLOMBIA

Introducción a la Minería de Datos

Verónica Guarín Escudero
Escuela de Estadística

Correo: jvguarine@unal.edu.co

Temas

- Esta presentación es una breve introducción a los sistemas de recomendaciones.
- Definición del problema de recomendación.
- Clasificación de estrategias.
- Foco en las reglas de asociación.

Sistemas de Recomendación

Dominios de Recomendación

U: Usuarios

I: Ítems

Películas

Música

Productos

Viajes

Libros

Citas

Profesionales

Amigos

Matriz de Utilidad

- Conjunto de Ítems: I = {i1, i2, i3, i4, i5, i6}
- Conjunto de Usuarios: U = {u1, u2, u3, u4}
- Los valores representan grados de preferencias de los usuarios sobre los ítems.

Usuarios \ Ítems	*i1*	*i2*	*i3*	*i4*	*i5*	*i6*
u1	?	1	?	5	?	?
u2	4	?	2	3	?	?
u3	?	4	?	?	5	3
u4	?	?	5	?	3	?

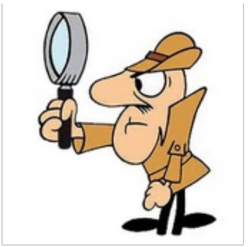
Representaciones

Usuarios \ Ítems	i1	i2	i3	i4	i5	i6
u1	?	1	?	5	?	?
u2	4	?	2	3	?	?
u3	?	4	?	?	5	3
u4	?	?	5	?	3	?

Explícita

Usuarios \ Ítems	i1	i2	i3	i4	i5	i6
u1	0	1	0	1	0	0
u2	1	0	1	1	0	0
u3	0	1	0	0	1	1
u4	0	0	1	0	1	0

Implícita



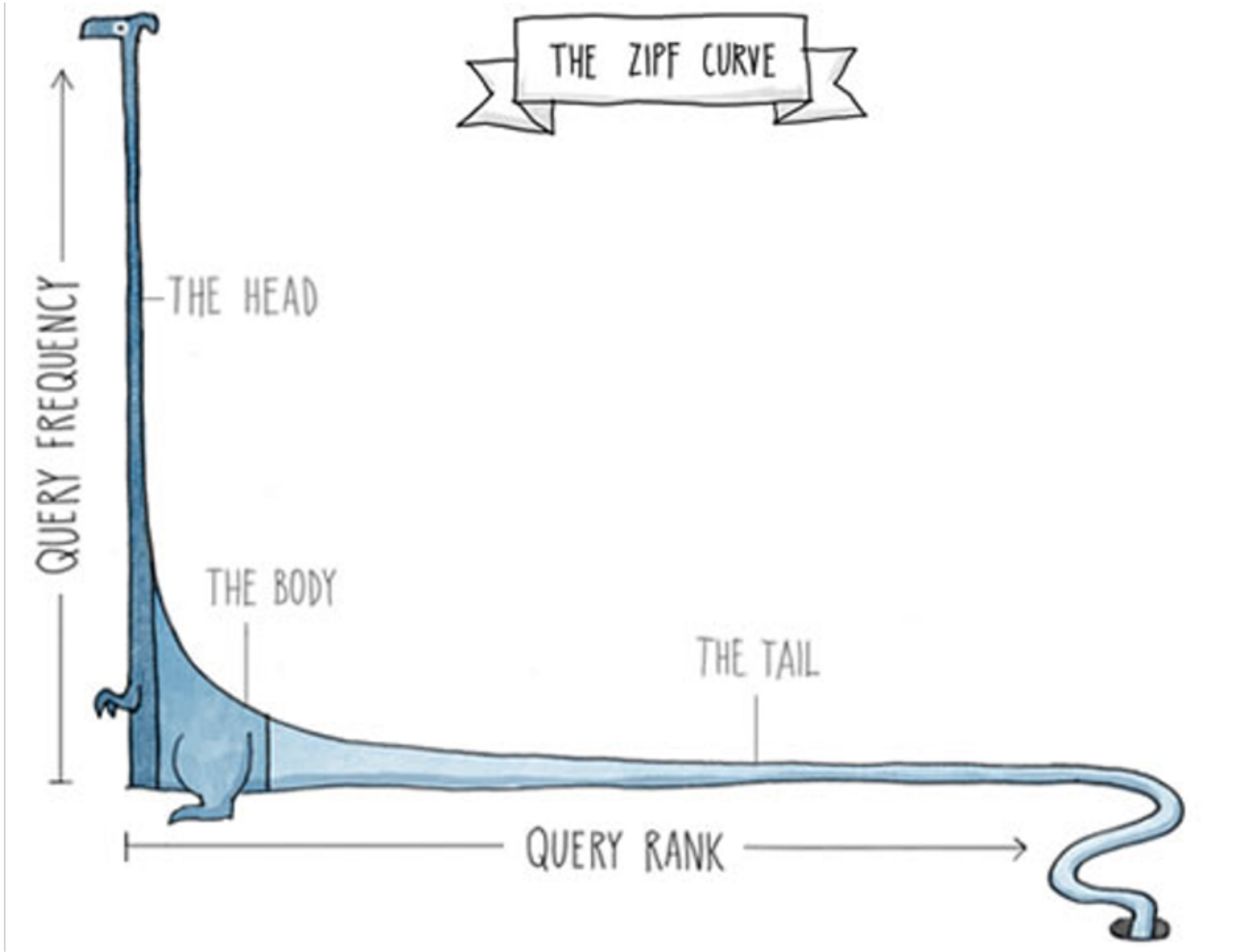
Etapas de recomendación

●Recomendaciones para el usuario: ***u***

Etapa 1: Predicción
●El sistema de recomendación asigna un score (valor numérico) a cada ítem ***i*** desconocido por ***u***.

Etapa 2: Recomendación
●Se genera una lista de ítems ordenada por valor de score, y se recomienda los primeros ***k*** elementos de esta lista.

Distribución de Cola larga



Clasificación de Algoritmos

- Popularidad
- Basados en contenido
- Asociación de productos
- Filtrado Colaborativo
- Híbridos y ensambles

Asociación de Productos

Los clientes que vieron este producto también vieron

Apple iPhone 7 - Smartphone de 4.7" (32 GB) oro

59

467,00 € prime

Apple iPhone XS (de 64GB) - Plata

7

1.066,39 € prime

Apple iPhone 8 Plus - Smartphone de 5.5" (64 GB) oro

26

739,00 € prime

Apple iPhone 7 Smartphone Libre Oro Rosa 128GB (Reacondicionado)

469

293,90 € prime

Apple iPhone 8 - Smartphone de 4.7" (256 GB) gris espacial

93

819,00 € prime

Reglas de Asociación: 2-itemsets

- Cada usuario es una transacción.
- Se calculan reglas entre todos los pares de ítems.
- Se utiliza una única métrica de asociación. (soporte, lift, etc)
- Se genera una matriz cuadrada **S** de tamaño |I| x |I|.
- Si la métrica utilizada es un índice de similitud. La matriz es simétrica, se la denomina matriz de similitud

Ítems		i1	i2	i3	i4	i5	i6
Ítems	i1	1	.2	.5	.1	.2	.9
	i2	.2	1	.05	.2	.4	0
	i3	.5	.05	1	.6	.2	0
	i4	.1	.2	.6	1	0	.9
	i5	.2	.4	.2	0	1	.7
	i6	.9	0	0	.9	.7	1

Reglas de Asociación: 2-itemsets

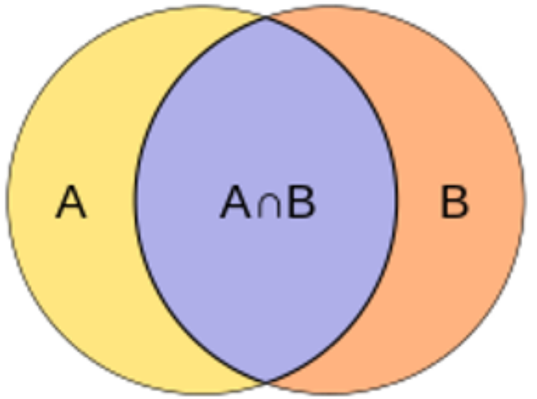
●Ejemplo de recomendación: El usuario ***u*** interactúa con el ítem ***i4***.

●Predicción
score(i1) = .1
score(i2) = .2
score(i3) = .6
score(i5) = 0
score(i6) = .9

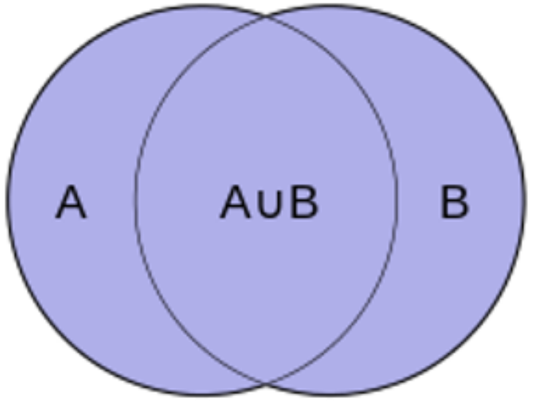
●Orden recomendación: [i6, i3, i2, i1, i5]

Ítems	i1	i2	i3	i4	i5	i6
i1	1	.2	.5	.1	.2	.9
i2	.2	1	.05	.2	.4	0
i3	.5	.05	1	.6	.2	0
i4	.1	.2	.6	1	0	.9
i5	.2	.4	.2	0	1	.7
i6	.9	0	0	.9	.7	1

Similitud Jaccard



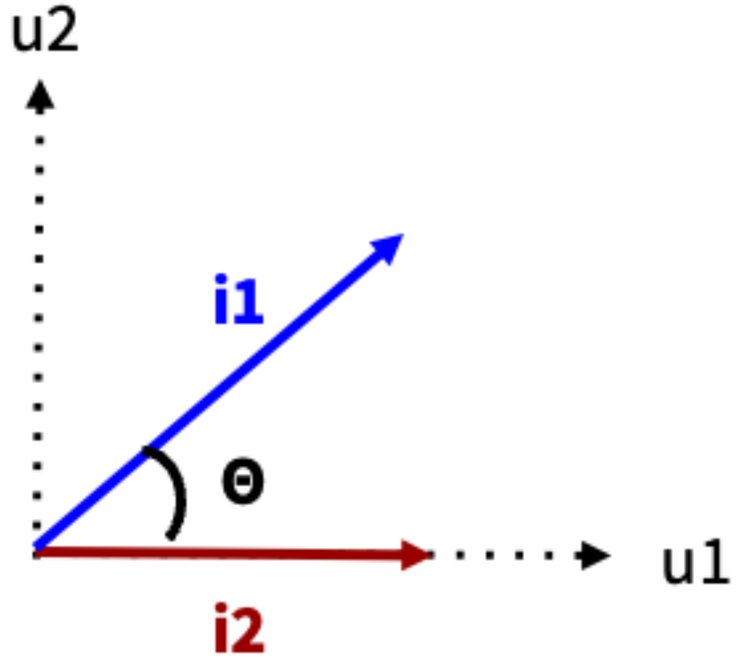
- Si los ítems se representan por conjuntos de usuarios.
- i1 = {u1, u2, u5}
- i2 = {u1, u6}
- sim(A, B) = supp(A & B) / (supp(A) + supp(B) - supp(A & B))
- sim(i1, i2) = |{u1}| / |{u1, u2, u5, u6}|
- sim(i1, i2) = 1/4 = .25



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

Similitud Coseno

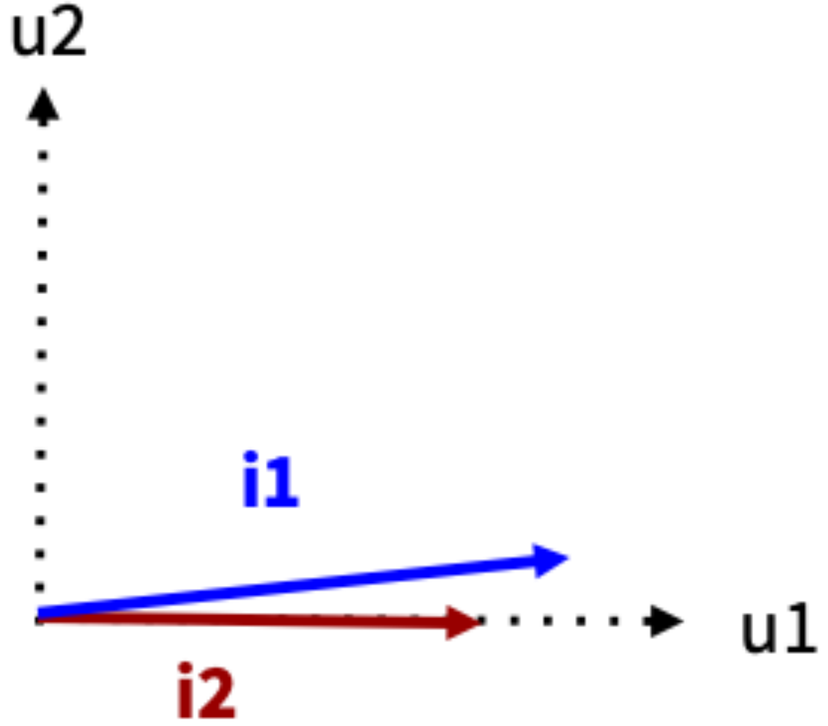
- Se representa a los ítems como vectores con tantas dimensiones como usuarios existan.
- i1 = {u1, u2, u5} -> [1,1,0,0,1,0]
- i2 = {u1, u6} -> [1,0,0,0,0,1]
- sim(X, Y) = supp(X & Y) / sqrt(supp(X) * supp(Y))
- sim(i1, i2) = 1/ raiz(2*3) = .41



$$similarity(A,B) = \frac{A \cdot B}{\|A\| \times \|B\|} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \times \sqrt{\sum_{i=1}^{n} B_i^2}}$$

Ejemplo Coseno

- i1 = {u1, u2, u5} -> [10,1,0,0,3,0]
- i2 = {u1, u6} -> [7,0,0,0,0,2]

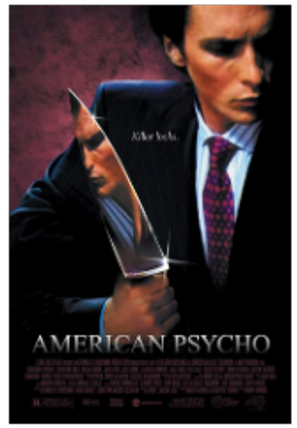


- **cos(i1,i2)** = (10*7) / (sqrt(10^2 + 1^2 + 3^2) * sqrt(7^2 + 2^2))
= **0.88**

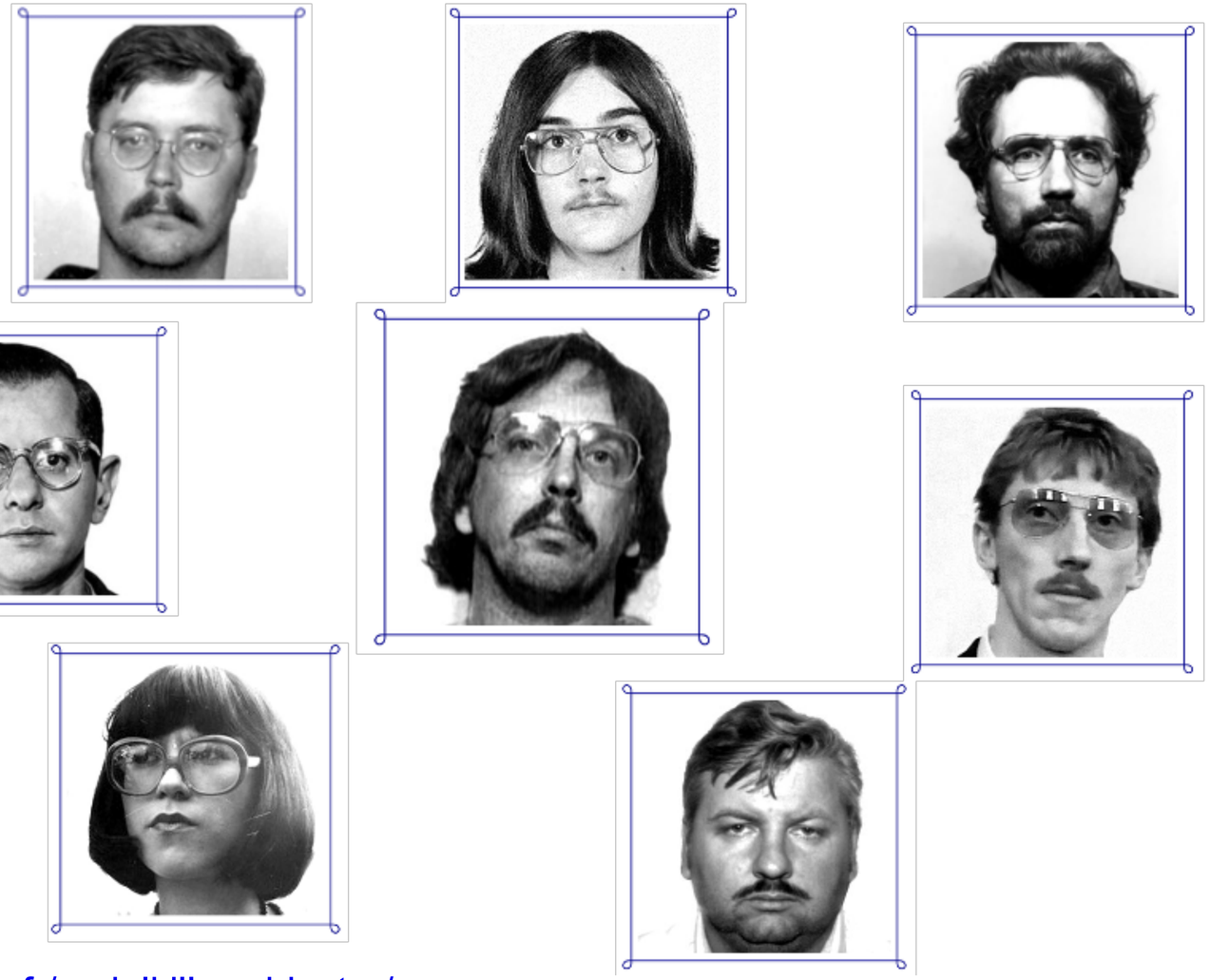
Filtrado Colaborativo

Joel

Ed



Filtrado Colaborativo: Users to users



http://noussommesbobbywatson.fr/serialkillerorhipster/

Filtrado Colaborativo: Users to users

Predicciones:

- Dado un usuario **u**
- Se calculan similitudes contra todos los usuarios.
- Se eligen a los *k* vecinos más cercanos.
- Por cada ítem ***i*** desconocido por ***u*** se calcula como **score** a la suma de todos los índices de similitud de los *k* vecinos que contienen a ***i***.

Filtrado Colaborativo: Items to items

- Es una generalización de las recomendaciones de asociación de productos.
- Para las predicciones se suman todas las similitudes a los ítems conocidos por ***u***.
- Ejemplo:

U1 = {i1, i3, i4}

score(ij) = sim(ij, i1) + sim(ij, i3) + sim(ij, i4)

score(i2) = .2 + .05 + .2 = 0.45

score(i5) = .2 + .2 = 0.4

score(i6) = .9 + .9 = 1.8

Ítems	i1	i2	i3	i4	i5	i6
i1	1	.2	.5	.1	.2	.9
i2	.2	1	.05	.2	.4	0
i3	.5	.05	1	.6	.2	0
i4	.1	.2	.6	1	0	.9
i5	.2	.4	.2	0	1	.7
i6	.9	0	0	.9	.7	1

Filtrado Colaborativo: Factorización de matrices

- Método de reducción de dimensionalidad.
  - SVD (singular vector descomposition)
  - Gradiente descendente
- Se descompone la matriz de utilidad en las nuevas dimensiones.
- Estas dimensiones latentes u ocultas captan distintas características de los ítems o usuarios.
- Los ítems y los usuarios quedan representados en este espacio latente.
- Al estar representados en un mismo espacio pueden calcularse directamente la similitud (o distancia) entre un usuario y un ítem.

Bibliografía

❑❑Jure Leskovec, Anand Rajaraman, Jeffrey D. Ullman (2014). Mining of Massive Datasets. Cambridge University Press. Segunda Edición. Capítulo 9 [http://www.mmds.org/]

❑Recommender Systems Specialization. University of Minnesota. Joseph A Konstan, Michael D. Ekstrand [https://www.coursera.org/specializations/recommender-systems]

UNIVERSIDAD
NACIONAL
DE COLOMBIA

Gracias