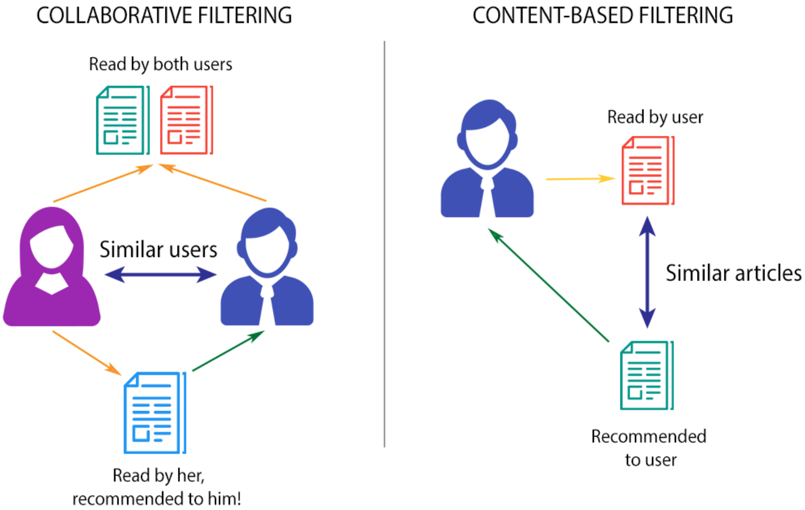
Factorization Machines

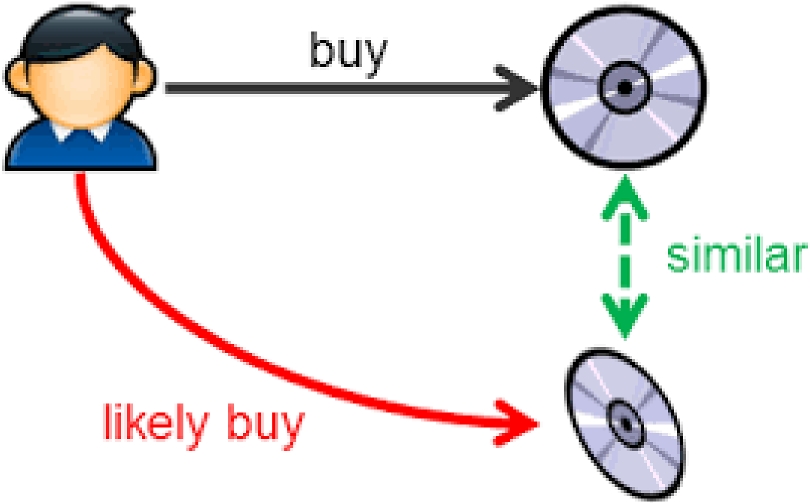
Sebastian Prillo

UBA - Seminario de Machin Lenin

Content-Based Filtering vs Collaborative Filtering



Content-Based Filtering: Ejemplo



Content-Based Filtering: Ejemplo

Content-Based Filtering: Ejemplo

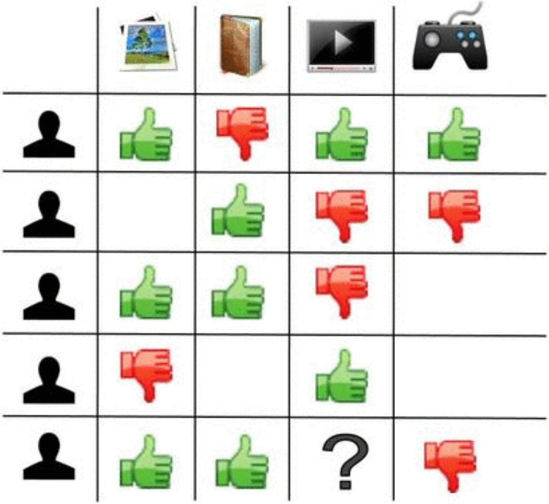
$$S(O_i, O_j) = \omega_1 f(A_{1i}, A_{1j}) + \omega_2 f(A_{2i}, A_{2j}) + \cdots + \omega_n f(A_{ni}, A_{nj})$$

Content-Based Filtering: Limitaciones

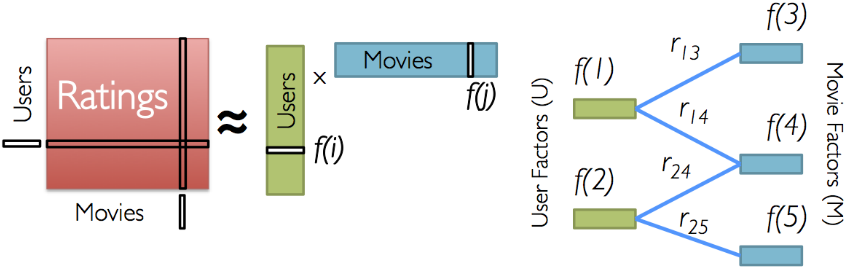
- Las representaciones de los items las construimos a mano.
- Que pasa si tenemos items de categorias distintas? ej: Televisores, libros, computadoras, ... Como defino su similitud?

Collaborative Filtering resuelve estos problemas (pero necesita mucha mas data)

Collaborative Filtering



Low-Rank Matrix Factorization:



Iterate:

$$f[i] = \arg\min_{w \in \mathbb{R}^d} \sum_{j \in \mathrm{Nbrs}(i)} \left(r_{ij} - w^T f[j]\right)^2 + \lambda ||w||_2^2$$

Content-Based Filtering: Limitaciones (Revisited)

- ~~Las representaciones de los items las construimos a mano.~~
  $\Rightarrow$ MF aprende solo las representaciones de los items y usarios! (son los vectores latentes que encuentra)
- ~~Que pasa si tenemos items de categorias distintas? ej: Televisores, libros, computadoras, ... Como defino su similitud?~~
  $\Rightarrow$ Las representaciones son compatibles!

$\Rightarrow$ TIRE TODA MI METADATA $\Leftarrow$

	Feature vector **x**																					Target y	
	User				Movie					Other Movies rated					Time	Last Movie rated							
	A	B	C	...	TI	NH	SW	ST	...	TI	NH	SW	ST	...		TI	NH	SW	ST	...			
$\mathbf{x}^{(1)}$	1	0	0	...	1	0	0	0	...	0.3	0.3	0.3	0	...	13	0	0	0	0	...		5	$y^{(1)}$
$\mathbf{x}^{(2)}$	1	0	0	...	0	1	0	0	...	0.3	0.3	0.3	0	...	14	1	0	0	0	...		3	$y^{(2)}$
$\mathbf{x}^{(3)}$	1	0	0	...	0	0	1	0	...	0.3	0.3	0.3	0	...	16	0	1	0	0	...		1	$y^{(2)}$
$\mathbf{x}^{(4)}$	0	1	0	...	0	0	1	0	...	0	0	0.5	0.5	...	5	0	0	0	0	...		4	$y^{(3)}$
$\mathbf{x}^{(5)}$	0	1	0	...	0	0	0	1	...	0	0	0.5	0.5	...	8	0	0	1	0	...		5	$y^{(4)}$
$\mathbf{x}^{(6)}$	0	0	1	...	1	0	0	0	...	0.5	0	0.5	0	...	9	0	0	0	0	...		1	$y^{(5)}$
$\mathbf{x}^{(7)}$	0	0	1	...	0	0	1	0	...	0.5	0	0.5	0	...	12	1	0	0	0	...		5	$y^{(6)}$

$$\hat{y}(x) = w_0 + \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n} \sum_{j=i+1}^{n} \langle v_i, v_j \rangle x_i x_j$$

- Anda sobre cualquier dataset, no solo user-item.
- Cuando tengo solo user-item, es equivalente a MF.

Factorization Machines

Una forma elemental de pensarlo: FMs = modelo lineal + k hyperplanos al cuadrado.

$$\hat{y}(x) = w_0 + \sum_{i=1}^{n} w_i x_i + \frac{1}{2} \sum_{f=1}^{k} \left(\left(\sum_{i=1}^{n} v_{i,f} x_i \right)^2 - \sum_{i=1}^{n} v_{i,f}^2 x_i^2 \right)$$

O sea, FMs = un modelo lineal al que le sumamos una no-linealidad muy sencilla.

DEMO