

Asignatura	
Técnicas de Inteligencia Artificial	Claudia Villalonga Palliser

## Actividades resueltas

### Filtrado colaborativo basado en ítems

#### Descripción de la actividad

Una tienda online tiene la siguiente base de datos que contiene información de las compras efectuadas por parte de cinco clientes.

	lector _ tarjeta	memoria _ usb	disco _ duro	pc _ sobremesa	portatil	pantalla	cargador _ universal	cd	teclado	dvd
<b>María</b>	0	1	0	1	0	1	0	1	1	1
<b>Juan</b>	1	1	0	1	0	1	0	1	0	1
<b>Rosa</b>	0	1	1	0	0	0	0	0	1	1
<b>Pedro</b>	0	1	0	1	1	0	1	0	1	0
<b>Ana</b>	1	1	0	0	0	1	1	0	0	1

Un valor de 1 para un determinado atributo quiere decir que el producto correspondiente ha sido adquirido por el cliente, mientras que un valor de 0 indica lo contrario, es decir que el producto no ha sido adquirido. Por ejemplo, María ha comprado en la tienda online una memoria USB, un PC de sobremesa, una pantalla, un CD, un teclado y un DVD.

Si Rosa está navegando por la web de la tienda y accede a la descripción del ítem 'portatil' **¿qué ítem se le recomendaría?** y si en lugar de recomendarle un único ítem se le recomendarán varios ítems ¿cuáles serían estos ítems y en qué orden se le recomendarían? Aplica el **algoritmo de filtrado colaborativo ítem-a-ítem** para identificar los productos que se le recomendarían. Describe claramente los cálculos efectuados por el algoritmo para medir la similitud entre ítems.

*Nota: No es necesario que calcules la matriz de similitud completa, puedes calcular solamente la similitud del ítem en cuestión con respecto al resto de ítems.*

Asignatura	
Técnicas de Inteligencia Artificial	Claudia Villalonga Palliser

## Resolución de la actividad

Para poder determinar los productos que un sistema de recomendación basado en el algoritmo de filtrado colaborativo ítem-a-ítem recomendaría a un Rosa cuando ésta accede a información del producto 'portatil', el primer paso es calcular la matriz de similitud de los ítems.

La matriz de similitud tiene una fila y una columna para cada uno de los ítems, es decir para cada uno de los productos registrados en la base de datos. Por tanto es una matriz cuadrada de dimensiones  $k \times k$ , donde  $k$  es el número de ítems (productos). En este caso la matriz de similitud tiene unas dimensiones de  $10 \times 10$ .

Cada elemento de la matriz de similitud es el valor de la similitud entre el ítem representado en la correspondiente fila y el ítem representado en la correspondiente columna. Este valor se mide en base a la medida de similitud coseno y se puede calcular con la siguiente fórmula:

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

donde cada ítem está representado por un vector  $\vec{A}$  (o  $\vec{B}$ ) de dimensión  $n$  igual al número de clientes de los que se tiene registros (5 en este caso).

La matriz de similitud es simétrica y en su diagonal tiene como valores unos porque la similitud de un ítem consigo mismo es máxima e igual a 1 en base a la fórmula anterior.

Tal como se indica en el enunciado no es necesario calcular la matriz completa sino que sólo la similitud del ítem 'portatil' (el producto que está consultando Rosa) con respecto al resto de ítems.

Realizando los cálculos, la similitud entre el ítem 'portatil' representado por el vector (0, 0, 0, 1, 0) y el ítem 'lector\_ tarjeta' representado por el vector (0, 1, 0, 0, 1) es:

$$similitud(portatil, lector\_tarjeta) = \frac{(0 \ 0 \ 0 \ 1 \ 0) \cdot (0 \ 1 \ 0 \ 0 \ 1)}{\|(0 \ 0 \ 0 \ 1 \ 0)\| \cdot \|(0 \ 1 \ 0 \ 0 \ 1)\|} = \frac{0}{\sqrt{1} \cdot \sqrt{2}} = 0$$

Asignatura	
Técnicas de Inteligencia Artificial	Claudia Villalonga Palliser

La similitud entre el ítem 'portatil' y el ítem 'memoria\_usb' es:

$$\text{similitud}(\text{portatil}, \text{memoria\_usb}) = \frac{(0\ 0\ 0\ 1\ 0) \cdot (1\ 1\ 1\ 1\ 1)}{\|(0\ 0\ 0\ 1\ 0)\| \cdot \|(1\ 1\ 1\ 1\ 1)\|} = \frac{1}{\sqrt{1} \cdot \sqrt{5}} = 0.447$$

La similitud entre el ítem 'portatil' y el ítem 'disco\_duro' es:

$$\text{similitud}(\text{portatil}, \text{disco\_duro}) = \frac{(0\ 0\ 0\ 1\ 0) \cdot (0\ 0\ 1\ 0\ 0)}{\|(0\ 0\ 0\ 1\ 0)\| \cdot \|(0\ 0\ 1\ 0\ 0)\|} = \frac{0}{\sqrt{1} \cdot \sqrt{1}} = 0$$

La similitud entre el ítem 'portatil' y el ítem 'pc\_sobremesa' es:

$$\text{similitud}(\text{portatil}, \text{pc\_sobremesa}) = \frac{(0\ 0\ 0\ 1\ 0) \cdot (1\ 1\ 0\ 1\ 0)}{\|(0\ 0\ 0\ 1\ 0)\| \cdot \|(1\ 1\ 0\ 1\ 0)\|} = \frac{1}{\sqrt{1} \cdot \sqrt{3}} = 0.577$$

La similitud entre el ítem 'portatil' y el ítem 'pantalla' es:

$$\text{similitud}(\text{portatil}, \text{pantalla}) = \frac{(0\ 0\ 0\ 1\ 0) \cdot (1\ 1\ 0\ 0\ 1)}{\|(0\ 0\ 0\ 1\ 0)\| \cdot \|(1\ 1\ 0\ 0\ 1)\|} = \frac{0}{\sqrt{1} \cdot \sqrt{3}} = 0$$

La similitud entre el ítem 'portatil' y el ítem 'cargador\_universal' es:

$$\text{similitud}(\text{portatil}, \text{cargador\_universal}) = \frac{(0\ 0\ 0\ 1\ 0) \cdot (0\ 0\ 0\ 1\ 1)}{\|(0\ 0\ 0\ 1\ 0)\| \cdot \|(0\ 0\ 0\ 1\ 1)\|} = \frac{1}{\sqrt{1} \cdot \sqrt{2}} = 0.707$$

La similitud entre el ítem 'portatil' y el ítem 'cd' es:

$$\text{similitud}(\text{portatil}, \text{cd}) = \frac{(0\ 0\ 0\ 1\ 0) \cdot (1\ 1\ 0\ 0\ 0)}{\|(0\ 0\ 0\ 1\ 0)\| \cdot \|(1\ 1\ 0\ 0\ 0)\|} = \frac{0}{\sqrt{1} \cdot \sqrt{2}} = 0$$

La similitud entre el ítem 'portatil' y el ítem 'teclado' es:

$$\text{similitud}(\text{portatil}, \text{teclado}) = \frac{(0\ 0\ 0\ 1\ 0) \cdot (1\ 0\ 1\ 1\ 0)}{\|(0\ 0\ 0\ 1\ 0)\| \cdot \|(1\ 0\ 1\ 1\ 0)\|} = \frac{1}{\sqrt{1} \cdot \sqrt{3}} = 0.577$$

La similitud entre el ítem 'portatil' y el ítem 'dvd' es:

$$\text{similitud}(\text{portatil}, \text{dvd}) = \frac{(0\ 0\ 0\ 1\ 0) \cdot (1\ 1\ 1\ 0\ 1)}{\|(0\ 0\ 0\ 1\ 0)\| \cdot \|(1\ 1\ 1\ 0\ 1)\|} = \frac{0}{\sqrt{1} \cdot \sqrt{4}} = 0$$

Los valores de similitud que se acaban de calcular se corresponden con las posiciones de la fila y la columna 'portatil' en la matriz de similitud. Por tanto la matriz de similitud sería tal que así:

Asignatura	
Técnicas de Inteligencia Artificial	Claudia Villalonga Palliser

	lector_ tarjeta	memoria_ usb	disco_ duro	pc_ sobremesa	portatil	pantalla	cargador_ universal	cd	teclado	dvd
lector_ tarjeta	1				0					
memoria_ usb		1			0.477					
disco_ duro			1		0					
pc_ sobremesa				1	0.577					
portatil	0	0.477	0	0.577	1	0	0.707	0	0.577	0
pantalla					0	1				
cargador_ universal					0.707		1			
cd					0			1		
teclado					0.577				1	
dvd					0					1

Una vez se han calculado las similitudes entre ítems se puede contestar a las preguntas. Si Rosa está navegando por la web de la tienda y accede a la descripción del ítem 'portatil' se le recomendaría al ítem 'cargador\_universal' porque es el que tiene mayor similitud, concretamente la medida de similitud del coseno es de 0.707. Si en lugar de recomendarle un único ítem se le recomendaran varios ítems, además del ítem 'cargador\_universal' el sistema recomendador le sugeriría a Rosa que comprar el ítem 'pc\_sobremesa', el ítem 'teclado' y el ítem 'memoria\_usb'. Lo haría en este orden porque va de mayor a menor similitud.