

# ¿La información es poder?

por

BEATRIZ RUBIO SERRANO

(Colaboradora del Instituto Universitario de Matemáticas y Aplicaciones)

Facebook está a punto de alcanzar los 1500 millones de usuarios, y no solo eso, aspira a que su aplicación esté operativa en los 7000 millones de teléfonos inteligentes que hay en el mundo. Amazon, el gigante del comercio internacional, se considera a sí misma una empresa de prestación de servicios, no una simple vendedora de productos. El servicio diferente que ofrece Amazon es poner en contacto las compras que hacemos online con el mundo real y físico. Amazon se ocupa de todo lo que sucede desde que haces clic y compras un artículo hasta que te llega a tu casa —ahora ya hasta mediante drones—, además del servicio de posventa.

Google —empresa creada por dos matemáticos— controla alrededor de 1,2 billones de búsquedas en Internet. La gran mayoría de las empresas se pelea por estar en las primeras posiciones del buscador más famoso. A través de Algebra Lineal y Grafos, el algoritmo google elige las páginas —PageRank— que ocuparán las primeras posiciones en nuestra búsqueda. Google es la agencia de publicidad más grande del planeta y la mayor base de datos privado de los que, como sabe interpretar, sabe monetizar.

Con la llegada de Internet y las nuevas tecnologías, cada uno de nosotros dejamos rastro, día a día, de una gran cantidad de datos: nuestros gustos, secretos, amigos..etc. Estos datos sí que son reales; son nuestras búsquedas, nuestras compras, nuestros amigos en facebook, nuestros «me gusta». Estos datos, sí que son reales, no son sacados de esas encuestas en las que los consumidores mentimos por naturaleza; estos datos son obtenidos de nuestros actos.

Hoy por hoy, las empresas tienen más información de nosotros pero por contra, nosotros — como consumidores— también tenemos más información de los productos que nos ofrecen. En pocos años hemos pasado de la escasez de información a la saturación. De disponer de algunas estanterías con CDs y Vídeos en la tienda de nuestro barrio se ha pasado a tener acceso a una cantidad inagotable de creaciones culturales en tiendas online o en redes P2P.

Ahora el problema para las empresas está en cómo separar lo que queremos de lo que no queremos encontrar y aquí es donde entran las matemáticas. Una aproximación para intentar ofrecer a cada persona lo que busca se consigue mediante los sistemas de recomendación.

Un sistema de recomendación es un tipo específico de filtro de información, técnica que trata de presentar al usuario items de información (películas, música, libros, noticias, páginas web) sobre las que el usuario está interesado. Para crear dicha base de datos se registran interacciones entre usuarios y artículos. No todos los sistemas de recomendación son iguales, nos centraremos en los de filtrado colaborativo.

Los sistemas de recomendación mediante algoritmos de filtrado colaborativo usan la información conocida sobre las preferencias de otros usuarios para realizar la recomendación al usuario que la precise. Lo que pretenden es identificar a aquellos usuarios cuyas preferencias sean similares a las de otros usuarios dados y así recomendar a los primeros los elementos que hayan satisfecho a los segundos. De esta forma, si dos usuarios  $U_1$  y  $U_2$  tienen las mismas preferencias y al usuario  $U_1$  le ha satisfecho el ítem  $i$ , probablemente este ítem también le satisfaga al usuario  $U_2$  por lo que también deberíamos recomendárselo. Esta situación puede ser representada como una matriz de usuarios por ítems, donde el elemento  $(i,j)$  representa la valoración de un usuario  $i$  con respecto a un ítem  $j$ .

Así visto, el algoritmo consiste en *predecir* que es lo que un consumidor haría. Y es que el filtrado colaborativo es un aspecto de gran importancia dentro de las redes sociales y la pequeña revolución que ha supuesto la llamada *Web 2.0*; Recomendaciones en tiendas online, filtrado de noticias, recomendaciones musicales, de libros, de películas, búsqueda de personas afines en comunidades y un largo etcétera.

Para determinar la similitud entre dos usuarios,  $w(a, u)$ , uno de los mecanismos más utilizado es el coeficiente de correlación de Pearson, donde la correlación entre los usuarios  $a$  y  $u$  se calcula como se muestra. En esta fórmula, el índice  $i$  corresponde a los ítems que fueron evaluados por los usuarios  $a$  y  $u$ .

$$w_{a,u} = \frac{\sum_{i=1}^m (r_{a,i} - \bar{r}_a) \cdot (r_{u,i} - \bar{r}_u)}{\sigma_a \cdot \sigma_u}$$

Donde  $i$  indica cuales son aquellos elementos para los que ambos usuarios han registrado votos y  $r_{a,i}$  representa la valoración del usuario  $a$  del ítem  $i$ .

En la siguiente tabla podemos ver un ejemplo de matriz que representa valoraciones de usuarios con respecto a una serie de películas.

	<i>Gladiator</i>	<i>Pretty Woman</i>	<i>Matrix</i>	<i>El padrino</i>	<i>Casino</i>	Media puntuaciones
Juan	5	3	4	4	¿?	4
Felipe	3	1	2	3	3	2,4
Marta	4	3	4	3	5	3,8
Sara	3	3	1	5	4	3,2
Raquel	1	5	5	2	1	2,8

El primer paso es medir la similitud de todos los usuarios con respecto al usuario activo, en nuestro caso Juan. Para ello utilizaremos el coeficiente de correlación de Pearson —podríamos usar otras métricas como MSD (Minimum Square Difference) o la del Coseno—.

$W(\text{Juan, Felipe})$	0,8528
$W(\text{Juan, Marta})$	0,7071
$W(\text{Juan, Sara})$	0,000
$W(\text{Juan, Raquel})$	-0,7921

Una vez que se han obtenido las similitudes, el objetivo es predecir la puntuación que Juan daría a la película *Casino*.

$$(\text{Juan, Casino}) = \frac{3 \cdot 0,8528 + 5 \cdot 0,7071}{0,85 + 0,7071} = 3,9093$$

Para realizar esta predicción escogeremos los dos usuarios más similares a Juan: Felipe y Marta. Utilizaremos la media de las puntuaciones de estos dos usuarios teniendo en cuenta las similitudes con Juan.

Este ejemplo es una sencilla demostración del valor que daría un usuario a una película, dependiendo de la similitud que tiene con otros usuarios. Este proceso se puede repetir con cuantas películas queramos y así poder recomendar a Juan las películas que obtengan mayor valor en nuestra predicción. Amazon, por ejemplo, utiliza este algoritmo para recomendar a los usuarios películas, libros, CDs...

Los sistemas de filtrado colaborativo también se utilizan en otros campos como por ejemplo en el de recuperación de información; la similitud entre dos documentos se mide a menudo considerando cada documento como un vector de frecuencias de palabras y calculando el coseno del ángulo formado por los dos vectores de frecuencia. En este caso, si lo relacionamos con el ejemplo anterior, los títulos de películas asumen el papel de las palabras y los votos de los usuarios son las frecuencias de la palabra.

Dentro del campo de la informática y del marketing, los sistemas de recomendación son uno de los miles de ejemplos en los que están presentes las matemáticas. Ha llovido mucho desde que [Alan Turing](#) —matemático, considerado por muchos historiadores padre de la informática moderna— demostró en 1936 que las máquinas de Turing —dispositivos que trabajaban con símbolos abstractos— podían resolver cualquier problema matemático que se representara con un algoritmo, hasta la realidad de este mundo internauta de redes sociales, compras on line, búsquedas de información, etc., y en el que, como hemos visto, el buen análisis de nuestros datos es fundamental para las grandes compañías. ¿Entonces?, ¿la información es poder? Mi opinión es que sí, si sabes interpretar; sí, si sabes matemáticas.