Workshop:

## SISTEMAS DE RECOMENDACIÓN

PARA LA PERZONALIZACIÓN DE LA OFERTA









Vicerrector de Relaciones Internacionales en Universidad de Deusto.

Me puedes encontrar como:

@alrayon

alex.rayon.jerez@gmail.com

www.alexrayon.es

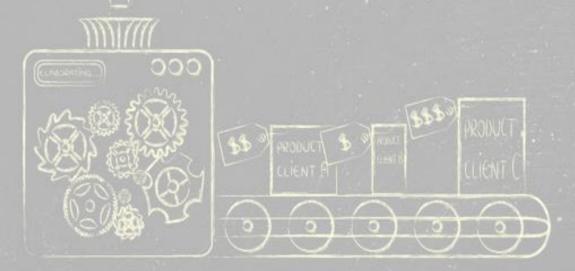
CUSTOMER NEEDS

PERSONAL PREFERENCE

INDIVIDUĂL STYLE

CUSTOM REQUESTS

# Hacia la economía de la experiencia



Relevancia vs. Fidelización

## Marketers Need to Stop Focusing on Loyalty and Start Thinking About Relevance

The problem, however, is that brands using the four P's exclusively often target a static customer archetype (e.g., a high-minded customer for an organic supermarket or a value-conscious customer for a discount chain). The reality is that there is no such archetypical customer. Everyone's needs vary depending on time and context. And with today's technologies, companies now have the ability to see and act on these fluctuations in the moment. Customers are increasingly expecting all companies to do just that, both in their marketing efforts and in the experiences they offer.

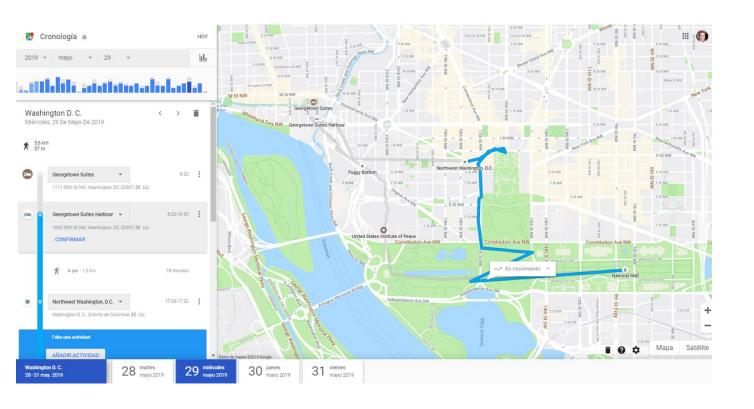
#### How Marketing Moved from the Mass-Market Era to the Era of Relevance

A combination of technology, data, and organizational goals has ushered in each new growth stage.

GROWTH ERA	MASS-MARKET	SEGMENT	CUSTOMER	LOYALTY	RELEVANCE
Decade	1960s-1970s	1980s	1990s	2010s	2020s
Technology enabler	Mass production	Market research	Enterprise IT	Advanced CRM	Digitization of everything
Performance indicator	Volume	Purchase funnel	Customer lifetime value	Customer retention	Customer attraction
Market approach	Mass appeal	Segmentation	Proposition innovation	Tailored incentives	Personalization
Management focus	Product and scale	Channel and scale	Channel and relationship	Experience and relationship	Experience and personality

SOURCE ACCENTURE © HBR.ORG

Dando una vuelta por Washington DC



Dando una vuelta por Washington DC (II)

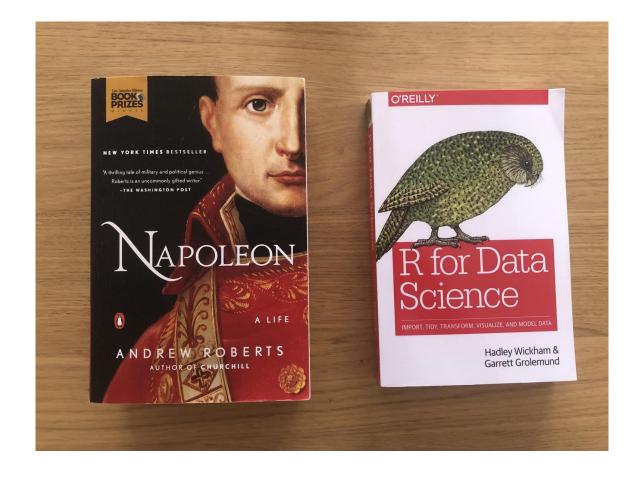


Dando una vuelta por Washington DC (III)





Dando una vuelta por Washington DC (IV)



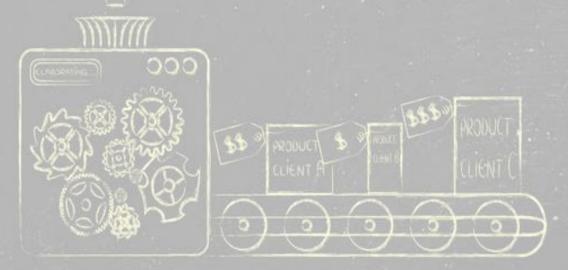
CUSTOMER NEEDS

PERSONAL PREFERENCE

INDIVIDUĂL STYLE

CUSTOM REQUESTS

# Sistemas de recomendación



¿Qué son?

- Herramientas diseñadas para interactuar con conjuntos de información grandes y complejos con la finalidad de proporcionar al usuario información o ítems que sean de su interés, todo ello de forma automatizada
- Se basa en el empleo de métodos matemáticos y estadísticos capaces de explotar la información previamente almacenada y crear recomendaciones adaptadas a cada usuario

¿Qué son? (II)

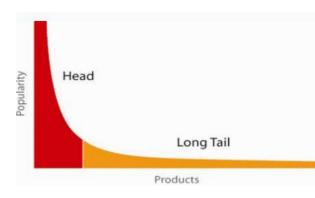
 Se puede entender como una función

#### Entrada

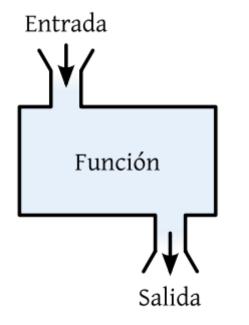
- Información de los usuarios (ej. puntuaciones (ratings), preferencias, variables demográficas, etc.)
- Objetos (ítems) (con o sin descriptores de sus características)

#### Salida

 Relevancia de un objeto → ranking de objetos



Recomendar objetos desconocidos que a los usuarios les pueda gustar



### ¿Por qué?

- Hoy en día, nos enfrentamos a un sinnúmero de alternativas al momento de escoger un producto o servicio
  - Tanto así, que muchas veces no logramos decidir con qué quedarnos porque quedamos sumidos entre tanta opción, una especie de parálisis del análisis
  - Nuestro cerebro parece querer optimizar la elección comparando los parámetros de cada uno, pero el espacio del problema es tan complejo que nos termina "fundiendo"
  - ¿Por qué lo hacemos? Tal vez sea para evitar arrepentirnos después, "mejor me compraba el otro"

¿Por qué? (II)

#### PERSONALITY PROCESSES AND INDIVIDUAL DIFFERENCES

## When Choice is Demotivating: Can One Desire Too Much of a Good Thing?

Sheena S. Iyengar Columbia University

Mark R. Lepper Stanford University

Current psychological theory and research affirm the positive affective and motivational consequences of having personal choice. These findings have led to the popular notion that the more choice, the better—that the human ability to manage, and the human desire for, choice is unlimited. Findings from 3 experimental studies starkly challenge this implicit assumption that having more choices is necessarily more intrinsically motivating than having fewer. These experiments, which were conducted in both field and laboratory settings, show that people are more likely to purchase gourmet jams or chocolates or to undertake optional class essay assignments when offered a limited array of 6 choices rather than a more extensive array of 24 or 30 choices. Moreover, participants actually reported greater subsequent satisfaction with their selections and wrote better essays when their original set of options had been limited. Implications for future research are discussed.

¿Por qué? (II)



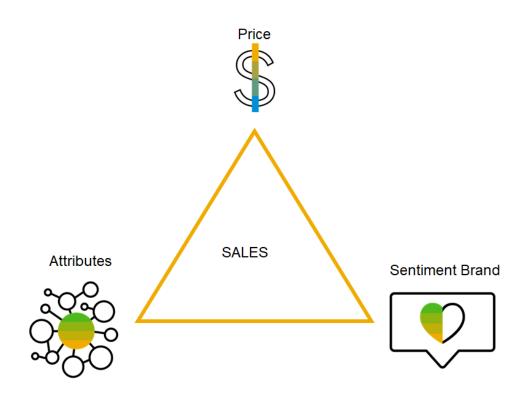
Front Psychol. 2015; 6: 1287. Published online 2015 Aug 28. doi: 10.3389/fpsyg.2015.01287 PMCID: PMC4551825

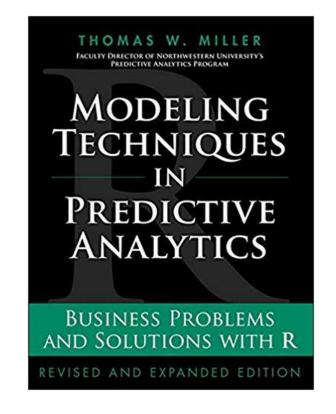
To make people save energy tell them what others do but also who they are: a preliminary study

Michele Graffeo, 1,\* Ilana Ritov, 2 Nicolao Bonini, 3 and Constantinos Hadjichristidis 3,4

"Se hizo la prueba yendo a un barrio hace unos años a ir a decirle a varios grupos de vecinos que tenían que consumir menos electricidad con argumentaciones diferentes. A uno se les explicó que era mejor para el medio ambiente. A otro que para ahorrar dinero. A un tercero que porque era bueno para la sociedad. Y al cuarto grupo le dijeron que era porque era lo que hacía la gran mayoría del barrio. Ganó claramente el último, porque el incentivo más poderoso es el social"

¿Cómo?





¿Quién lo hace?

Industry Report



Item-to-Item Collaborative Filtering

Greg Linden, Brent Smith, and Jeremy York • Amazon.com



#### ¿Qué analizan? (entre otras cosas?

- (1) Tus compras
- (2) Carritos abandonados
- (3) Visualización de productos

Your Recently Viewed Items and Featured Recommendations Inspired by your browsing history











Hook - Hand Held Massage Tool - Use With RumbleRoller \$24.95 \Prime



Massage Balls for Deep Tissue, Trigger Point & 常常常常常 237 \$14.97 Prime



Premium High Density Foam Roller - Extra Firm With 1 Year **常常常常** 1,421 \$8.75 - \$24.95

2 X LACROSSE BALLS FOR TRIGGER POINT MASSAGE- Fine-Toned@ plus MASSAGE 常常常常常 43

\$10.75 \Prime

Page 1 of 8

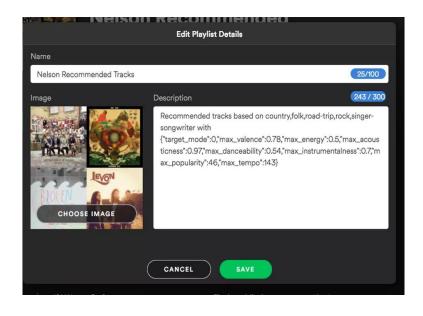
Deadlift Jack Alternative for Your Gym Bag Raises loaded barbell &

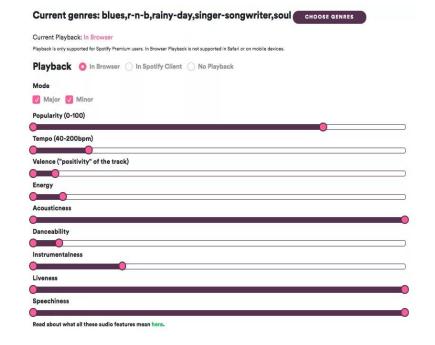
\$17.95 \Prime

¿Quién lo hace? (II)

## You can now play with Spotify's recommendation algorithm in your browser

It's fun if you know what 'valence' means!





¿Quién lo hace? (III)



HOME

RESEARCH AREAS >

BUSINESS AREAS ~

ADTICL ES

EVENTS & UPDATES

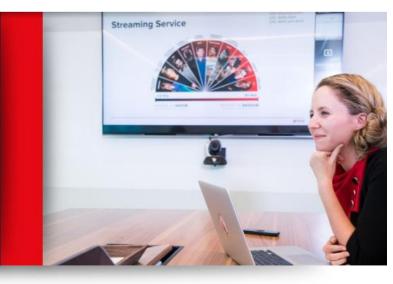
LODG

Q

RESEARCH AREAS

### Recommendations

Figuring out how to bring unique joy to each member



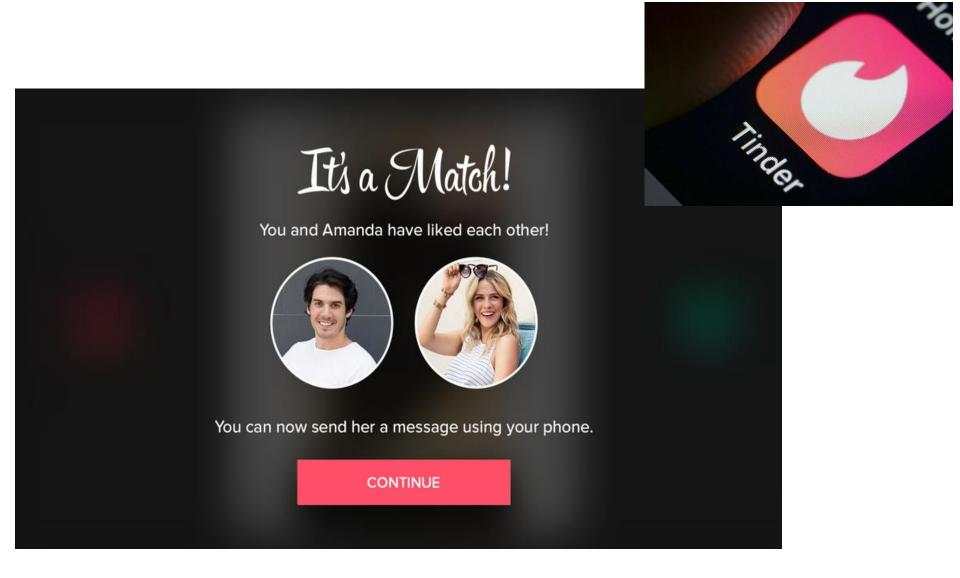
#### **About**

Recommendation algorithms have been the core of the Netflix product from very early on.

#### **Recent Job Openings**

Senior Research Scientist - Machine Learning

¿Quién lo hace? (IV)



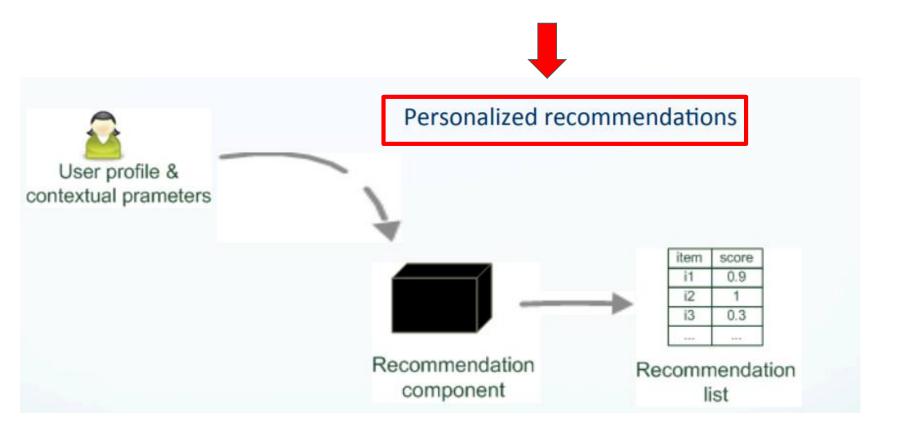
#### Marketing personalizado

#### Highly personalized marketing

Habits of consuming content have changed dramatically. US consumers doubled their spending on digital newspapers in the past seven years, for example, while halving their spending on print newspapers. [10] As more consumers abandon print media for digital media, marketers follow: 44 percent of them now allocate at least half of their marketing budgets to digital media, up from only 31 percent in 2009. [11]

We're already seeing that direct mail and newspaper circulars are playing a diminished role in retail marketing. Mass advertising will not disappear overnight, but its influence is certainly waning. Ads are shifting toward not just digitization but also personalization, powered by increasingly sophisticated algorithms and predictive models that analyze transaction data and digital-media trends (for example, what topics are hot on social networks). Already, 35 percent of what consumers purchase on Amazon and 75 percent of what they watch on Netflix come from product recommendations based on such algorithms.

Esquema de componentes

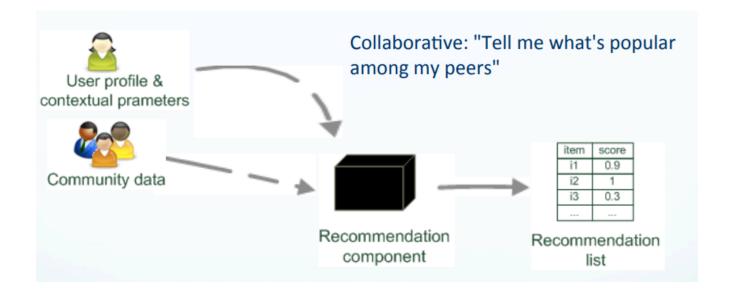


#### Considerando los datos usados

- Siguiendo a **Mobasher**, podemos establecer cuatro enfoques principales:
  - (1) Filtrado colaborativo
    - Las recomendaciones a un usuario se basan en las preferencias de "usuarios similares"
    - IBCF (Item-Based) y UBCF (User-Based)
  - (2) Filtrado basado en contenido
    - Las recomendaciones se basan en sus propias preferencias, buscando elementos similares por los que ha mostrado interés
  - (3) Filtrado basado en reglas
    - Las recomendaciones se basan en reglas previamente definidas como, por ejemplo, en características sociodemográficas
  - (4) Enfoques mixtos
    - Combinan diferentes enfoques



- (1) Filtrado colaborativo
  - Se basa en la premisa de que si tengo gustos en común con otro usuario (ratings similares a las cosas que ambos hemos consumido) es probable que también me gusten



(1) Filtrado colaborativo (II)

#### Ejemplo

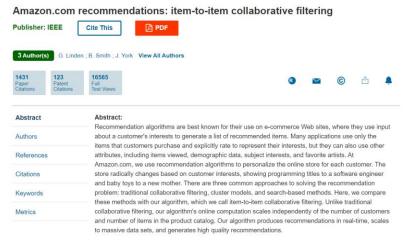
Set of users **U** and a set of items **I** to be recommended to the users.

Learn a function based on the past data that predicts utility of each item  $i \in I$  to each user  $u \in U$ 



Esta matriz se denomina **matriz de interacción** o **matriz de ratings** (en el "raro" caso que explícitamente un usuario dé un rating a un ítem). La tarea de Machine Learning es construir una función que trate de predecir esa utilidad de cada ítem para cada usuario. La matriz normalmente es grande y con muchos valores que falten (escasez).

- (1) Filtrado colaborativo: IBCF y UBCF
  - IBCF fue lanzado por Amazon.com en 1998 (paper en IEEE de 2003)
    - Incrementó dramáticamente la escalabilidad de los sistemas de recomendación para incluir millones de productos y clientes
    - Posteriormente, ha sido adoptado por los grandes players tecnológicos (Youtube, Netflix, etc.)
  - UBCF tiene más años
    - El problema radica en sus problemas para la escalabilidad, así como su coste computacional





#### **IBCF vs UBCF**

Name	Avengers	Star wars	Thor	Spider-man	Iron Man
Alex	4	2	?	5	4
Bob	5	3	4	?	3
Tom	3	?	4	4	3

#### **UBCF**

(1) Calcular la similaridad entre Alex y otros usuarios

$$sim(x,y) = \cos(\vec{x}, \vec{y}) = \frac{\vec{x} \cdot \vec{y}}{||\vec{x}||_2 \times ||\vec{y}||_2} = \frac{\sum\limits_{s \in S_{xy}} r_{x,s} r_{y,s}}{\sqrt{\sum\limits_{s \in S_{xy}} r_{x,s}^2} \sqrt{\sum\limits_{s \in S_{xy}} r_{y,s}^2}},$$
(13)

- (2) Predecir la fórmula de rating que ha venido aplicando Alex
- (3) Seleccionar las películas con mayor rating

Name	Avengers	Star wars	Thor	Spider-man	Iron Man
Alex	4	2	4.02	5	4

(a) 
$$r_{c,s} = \frac{1}{N} \sum_{d' \in \hat{G}} r_{d,s},$$
  
(b)  $r_{c,s} = k \sum_{d' \in \hat{G}} sim(c, d') \times r_{d,s},$   
(c)  $r_{c,s} = \bar{r}_c + k \sum_{d' \in \hat{G}} sim(c, d') \times (r_{d',s} - \bar{r}_{d'}),$  (10)

(c) 
$$r_{c,s} = \bar{r}_c + k \sum_{c' \in \hat{C}} sim(c, c') \times (r_{c',s} - \bar{r}_{c'}),$$

where multiplier k serves as a normalizing factor and is usually selected as  $k = 1/\sum_{c' \in \hat{C}} |sim(c, c')|$ , and where the average rating of user c,  $\bar{r}_c$ , in (10c) is defined as<sup>1</sup>

IBCF vs UBCF (II)

Name	Avengers	Star wars	Thor	Spider-man	Iron Man
Alex	4	2	?	5	4
Bob	5	3	4	?	3
Tom	3	?	4	4	3

#### IBCF

- (1) Transponer la matriz usuario-item a la matriz itemusuario
  - (2) Calcular la similitud entre dos ítems y completar la matriz de similitud ítem-ítem
  - (3) Predecir la fórmula de rating que ha venido aplicando Alex
  - (4) Seleccionar las películas con mayor rating

	Alex	Bob	Tom
Avengers	4	5	3
Star wars	2	3	?
Thor	?	4	4
Spider-man	5	?	4
Iron Man	4	3	3

$$sim(i,j) = \cos(\vec{i},\vec{j}) = \frac{\vec{i} \cdot \vec{j}}{||\vec{i}||_2 * ||\vec{j}||_2}$$

	Avengers	Star wars	Thor	Spider-man	Iron Man
Avengers					
Star wars	0.99624059				
Thor	0.9701425	0			
Spider-man	0.99951208	0	0		
Iron Man	0.9701425	0.94299033	1	0.99951208	

(2) Filtrado basado en contenido

• *Ejemplo*: Pandora

Una radio online que genera listas de reproducción adaptadas a cada usuario. Para lograrlo, dispone de una base de datos con una descripción detallada sobre el ritmo, estilo de música, instrumentos, temática... de cada canción. A partir de que el usuario escoge una canción inicial, se empiezan a sugerir canciones similares



(2) Filtrado basado en contenido (II)

- *Ejemplo*: Amazon
  - Un libro parecido a otro que se ha leído, por ejemplo, por pertenecer al mismo autor.
  - También podría ser del mismo género literario o ambientado en una determinada época histórica
  - De esta misma forma te anima Amazon a añadir a la cesta productos similares a aquéllos que estás buscando.

#### Recomendado para ti en Libros



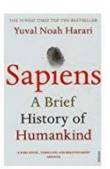




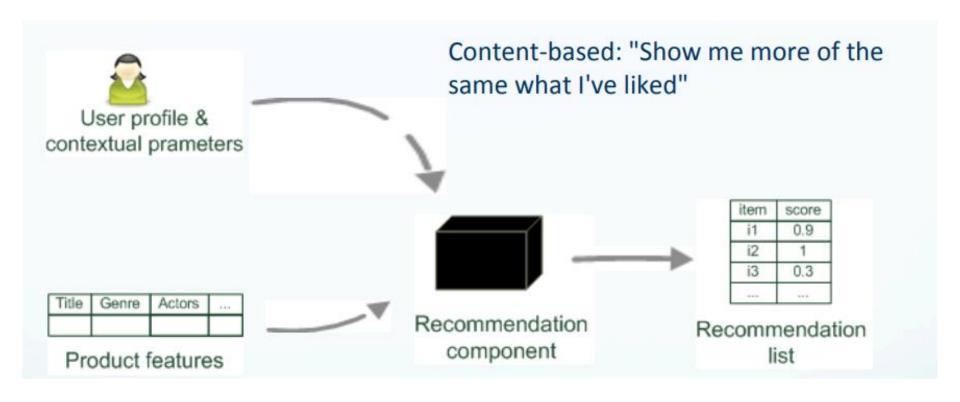




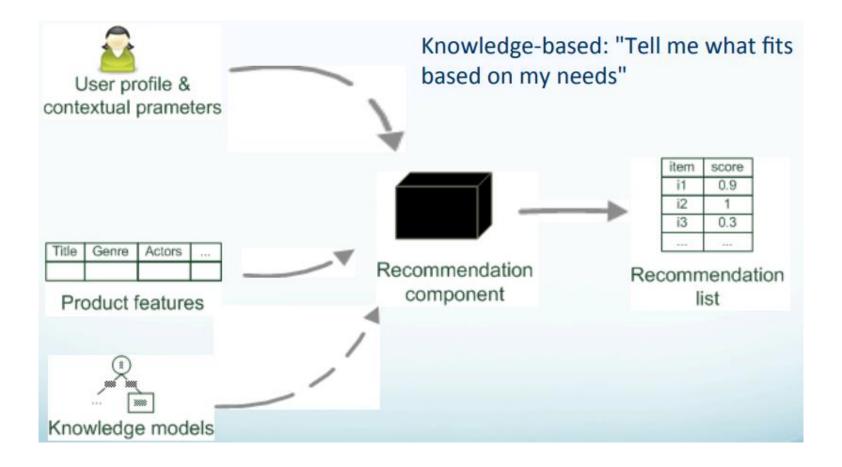




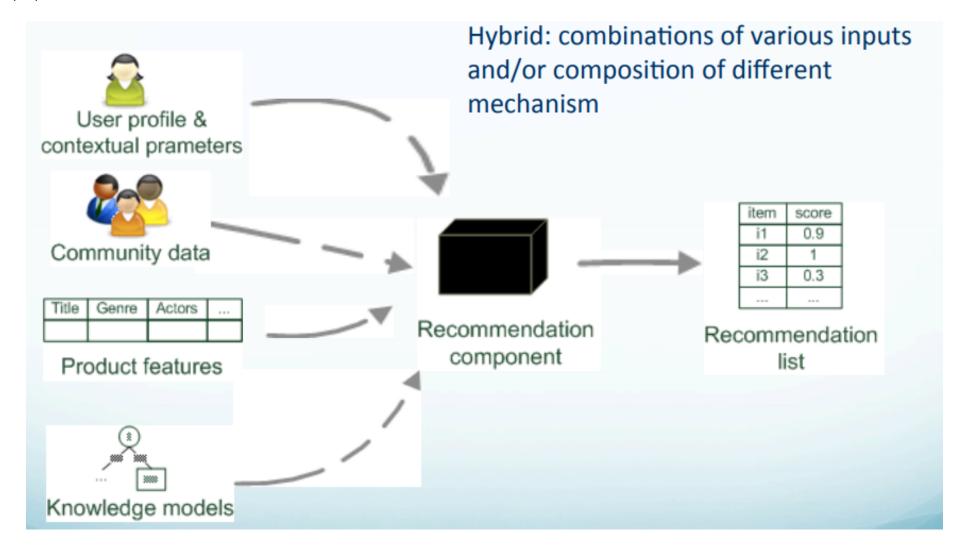
(2) Filtrado basado en contenido (III)



(3) Filtrado basado en reglas



(4) Mixtos



#### Medidas de similitud o distancia

- Todos los sistemas de recomendación tienen una cosa en común:
  - Para poder llevar a cabo las predicciones, necesitan definir y cuantificar la similitud entre ítems o usuarios
- El término distancia se emplea dentro del contexto de recomendadores como cuantificación de la similitud o diferencia entre observaciones
  - Si se representan las observaciones en un espacio p dimensional, siendo p
    el número de variables asociadas a cada observación (ítem o usuario),
    cuando más se asemejen dos observaciones, más próximas estarán
  - De ahí que se emplee el término distancia
    - La característica que hace de los sistemas de recomendación un método adaptable a escenarios muy diversos es que pueden emplear cualquier tipo de distancia, lo que permite al científico o científica de datos escoger la más adecuada para el estudio en cuestión.

Medidas de similitud o distancia (II)

- La idea principal es calcular similaridades entre usuarios o ítems
  - Ese cálculo se hará de manera predictiva
  - A su vez, la capacidad de predecir dependerá de la medida de similaridad empleada para encontrar usuarios parecidos
  - Hay tres familias de medidas de similaridad:
    - (a) Basadas en correlación
      - Correlación de Pearson, Correlación restringida de Pearson, Spearman, correlación τ de Kendall, etc.
    - **■** (b) Basadas el vector de cosenos
    - (c) Basadas en la probabilidad condicionada

#### Medidas de similitud o distancia (III)

- Medidas de similaridad
  - (1) Correlación de Pearson
    - Es la más utilizada. Se emplea para calcular la correlación lineal entre dos vectores.
  - (2) Correlación restringida de Pearson
    - Utilizada valores medianos en lugar de medias de rating de dos usuarios.
  - o (3) Coseno
    - Es otra de las más utilizadas. Trata de encontrar cómo dos vectores están relacionados entre ellos midiendo el coseno del ángulo entre los dos vectores. El único problema que presenta es que considera también las preferencias nulas.

#### Medidas de similitud o distancia (IV)

- Medidas de similaridad
  - (4) Jaccard
    - Considera el número de preferencias compartidas entre dos usuarios. Dos usuarios serán más parecidos cuando tengan valores informados (ratings informados) similares.
  - (5) Diferencias de medias cuadradas
    - Considera los ratings absolutos, en lugar de qué ítems han sido valorados. Normalmente se suelen combinar la anterior (Jaccard) y esta.
  - (6) PIP: *Proximity, Impact, Popularity* 
    - Proximity → diferencia aritmética entre dos rating.
    - Impact → si un ítem le gusta o no al usuario.
    - Popularity → media aritmética del rating de ítems del usuario.

#### Medidas de similitud o distancia (V)

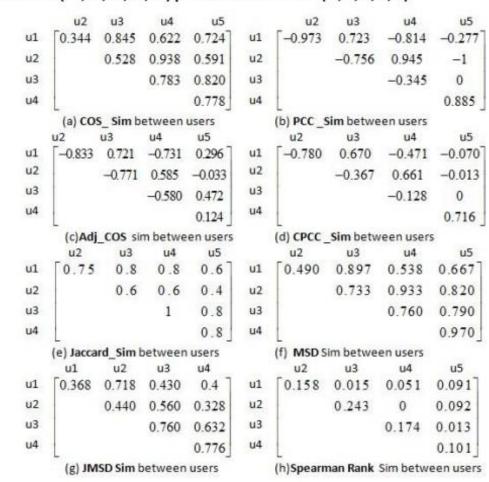
- Medidas de similaridad (II)
  - (7) New Heuristic Similarity Model (NHSM)
    - Entre 0 y 1, considera el hecho de que usuarios diferentes tienen diferentes escalas de preferencia y modela la preferencia de ellos en función de la media y la desviación típica del rating.
  - (8) Correlación de Spearman
    - Considera el ranking en lugar del rating para calcular la similitud.
  - (9) Correlación Tau de Kendall
    - Considera los rankings relativos en lugar del rating para calcular la similitud.

Medidas de similitud o distancia (VI)

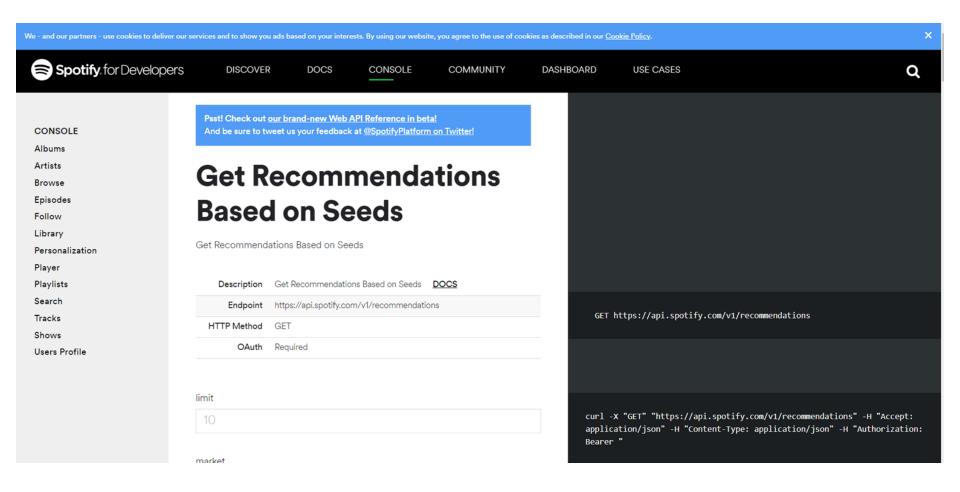
Figure 1: Matrix of users (u1,u2,u3,u4,u5) preferences on items (i1,i2,i3,i4,i5)

#### Matriz U, I

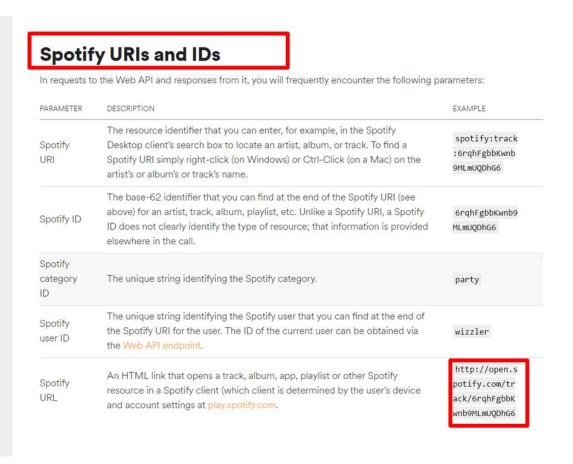
item	i1	i2	i3	i4	i5
user					
u1	5		0.5	1	5
u2	2		5	4	
u3	4	3	1	4	5
u4	4	1	.5	5	2
u5	5	1		4	1

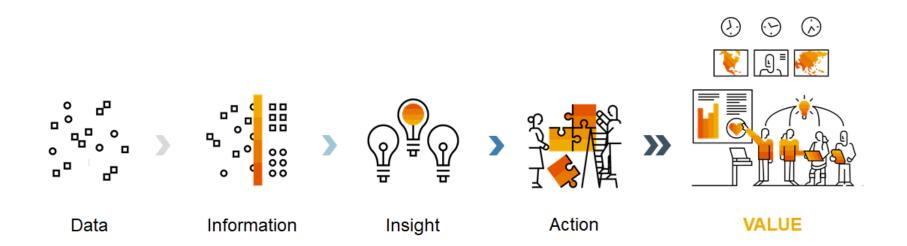


API Spotify: obtener recomendaciones



API Spotify: obtener recomendaciones (II)





# "Obsess for the customers, not for the competition"

Jeff Bezos - CEO Amazon





@alrayon



www.alexrayon.es



alex.rayon.jerez@gmail.com



www.linkedin.com/in/alrayon



@alrayon



http://youtube.com/c/alrayon



www.slideshare.net/alrayon



Workshop Sistemas de Recomendación

DMC Perú - DMC Online - #YoMeQuedoEnCasa 1

Lima, Marzo, 2020

Alex Rayón Jerez @alrayon, <u>alex.rayon.jerez@gmail.com</u> www.alexrayon.es