

ACÀMICA

¡Bienvenidas/os a Data Science!



Agenda

¿Cómo anduvieron?

Actividad: Data Science en mi vida

Repaso: Sistemas de recomendación

Break

Hands-on training

Cierre



¿Dónde estamos?



¿Cómo anduvieron?



Actividad: Data Science en mi vida



Data Science en mi vida

¡Preparen sus charlas relámpago!

En 7 minutos con 7 slides comparte con tus compañeros:

En qué problemas estás aplicando lo aprendido en Data Science y cómo lo estás haciendo.

O bien, en qué problemas te gustaría aplicar Data Science y cómo lo harías.

¡Elige algún tema o proyecto que te interese y relaciónalo con lo aprendido!

IMPORTANTE

Si no lo hicieron, bajar el siguiente dataset:

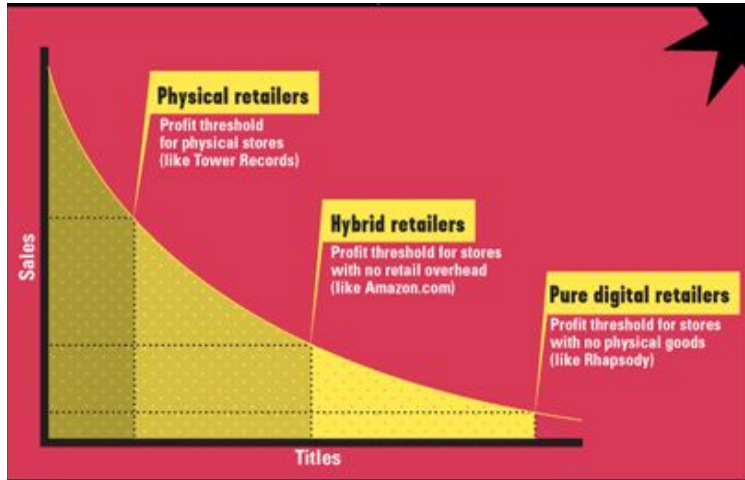
<https://www.kaggle.com/netflix-inc/netflix-prize-data>

Repaso: Sistemas de Recomendación



The long tail

Ejemplo: Into Thin Air y Touching Void



Matriz de utilidad

- Existen *usuarios* e *ítems*. Los *usuarios* prefieren algunos *ítems* por sobre otros
- Ejemplo: Usuarios de Netflix y Películas. De 1 a 5 estrellas.

	P1	P2	P3	P4	P5	P6	...	P _m
Usuario 1	5	4			2		...	1
Usuario 2	2	1		5			...	5
Usuario 3		1	5		4	3	...	2
Usuario 4	4			2	1			
...	
Usuario n	1	2	5		5		...	3

Matriz de utilidad

- Existen *usuarios* e *ítems*. Los *usuarios* prefieren algunos *ítems* por sobre otros
- Ejemplo: Usuarios de Netflix y Películas. De 1 a 5 estrellas.

	P1	P2	P3	P4	P5	P6	...	P _m
Usuario 1	5	4	?	?	2	?	...	1
Usuario 2	2	1	?	5	?	?	...	5
Usuario 3	?	1	5	?	4	3	...	2
Usuario 4	4	?	?	2	1	?	...	?
...
Usuario n	1	2	5	?	5	?	...	3

Matriz de utilidad

- Existen *usuarios* e *ítems*. Los *usuarios* prefieren algunos *ítems* por sobre otros
- Ejemplo: Usuarios de Netflix y Películas. De 1 a 5 estrellas.

El objetivo del sistema de recomendación es *poblar* la matriz de utilidad

	P1	P2	P3	P4	P5	P6	...	P _m
Usuario 1	5	4	?	?	2	?	...	1
Usuario 2	2	1	?	5	?	?	...	5
Usuario 3	?	1	5	?	4	3	...	2
Usuario 4	4	?	?	2	1	?	...	?
...
Usuario n	1	2	5	?	5	?	...	3

Matriz de utilidad

- Existen *usuarios* e *ítems*. Los *usuarios* prefieren algunos *ítems* por sobre otros
- Ejemplo: Usuarios de Netflix y Películas. De 1 a 5 estrellas.

Ejemplo: Netflix tiene 150 millones suscriptores, 5 mil películas. La matriz tiene 750000000000 espacios, de los cuales la mayoría están vacíos.

- Cuando buscamos recomendar, interesa más recomendar ítems que van a gustar que aquellos que no van a gustar.

Matriz de utilidad

- Existen *usuarios* e *ítems*. Los *usuarios* prefieren algunos *ítems* por sobre otros
- Ejemplo: Usuarios de Netflix y Películas. De 1 a 5 estrellas.

Ejemplo: Netflix tiene 150 millones suscriptores, 5 mil películas. La matriz tiene 750000000000 espacios, de los cuales la mayoría están vacíos.

- Cuando buscamos recomendar, interesa más recomendar ítems que van a gustar que aquellos que no van a gustar.
- En algunos casos, interesa mover a los usuarios del mainstream a la cola

Matriz de utilidad

- Existen *usuarios* e *ítems*. Los *usuarios* prefieren algunos *ítems* por sobre otros
- Ejemplo: Usuarios de Netflix y Películas. De 1 a 5 estrellas.

Ejemplo: Netflix tiene 150 millones suscriptores, 5 mil películas. La matriz tiene 750000000000 espacios, de los cuales la mayoría están vacíos.

- Cuando buscamos recomendar, interesa más recomendar ítems que van a gustar que aquellos que no van a gustar.
- En algunos casos, interesa mover a los usuarios del mainstream a la cola
- Algunas veces, ni siquiera hay calificaciones, solamente si vio o no (o escuchó, leyó, compró, etc.).

Matriz de utilidad

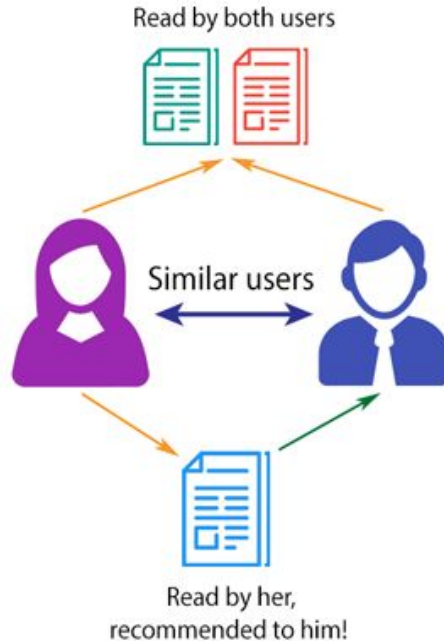
- Existen *usuarios* e *ítems*. Los *usuarios* prefieren algunos *ítems* por sobre otros
- Ejemplo: Usuarios de Netflix y Películas. De 1 a 5 estrellas.

Ejemplo: Netflix tiene 150 millones suscriptores, 5 mil películas. La matriz tiene 750000000000 espacios, de los cuales la mayoría están vacíos.

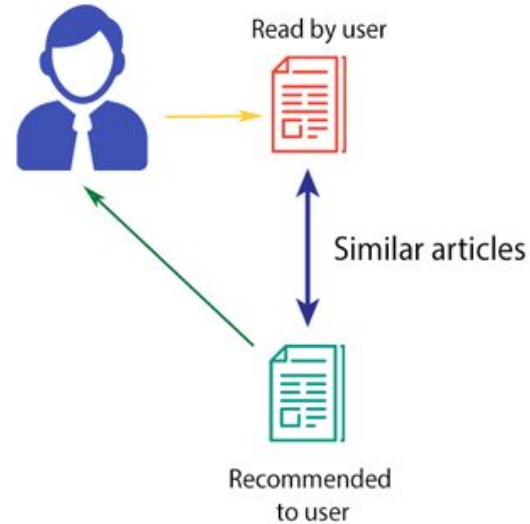
El objetivo del sistema de recomendación es poblar la matriz de utilidad de una manera inteligente y bajo los requisitos que imponga cada entorno

Tipos de Sistemas de Recomendación

COLLABORATIVE FILTERING



CONTENT-BASED FILTERING



Sistemas de recomendación • Tipos

	¿Cómo?	Ventaja	Desventaja / Problema
Basado en contenidos	Recomienda ítems con características similares a los que el usuario consumió (y, preferiblemente, indicó que le gustaban).	Basta con <i>conocer</i> los ítems para comenzar a recomendar	1) calcular la similitud entre dos ítems puede ser una tarea difícil y muy costosa. En la mayoría de los casos hay que obtener atributos. 2) Suele recomendar ítems que no son novedosos para el usuario
Filtro colaborativo	Recomienda ítems basadas en medidas de similaridad entre ítems y/o usuarios.	No necesita <i>conocer</i> los ítems, en principio alcanza con la información de la matriz de utilidad	Necesito la matriz de utilidad
Pensarlo como problema de clasificación	Podemos entrenar un clasificador para cada usuario		Pocas calificaciones por usuario
Híbridos	Combinar lo mejor de varios mundos		

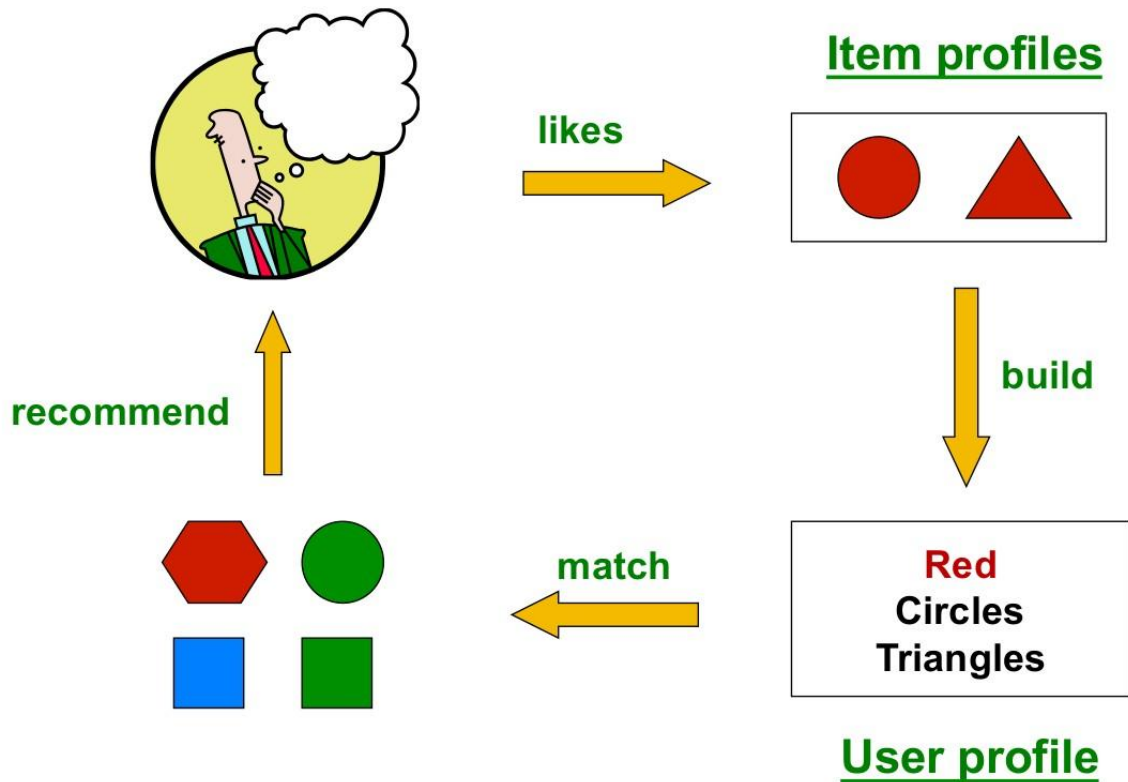
Sistemas de recomendación • Tipos

	¿Cómo?	Ventaja	Desventaja / Problema
Basado en contenidos	Recomienda ítems con características similares a los que el usuario consumió o expresó preferencia por.	Basta con <i>conocer</i> los ítems para comenzar a recomendar.	1) calcular la similitud entre dos ítems puede ser una tarea difícil y muy costosa. En la mayoría de los casos hay que calcular la similitud entre todos los ítems. 2) recomendar ítems que no son interesantes para el usuario.
Filtro colaborativo	Recomienda ítems basados en la similitud entre usuarios y/o usuarios e ítems.	matriz de utilidad	1) Necesidad de una gran cantidad de datos de utilidad. 2) Problema de la rareza de los ítems.
Pensarlo como problema de clasificación	Podemos entrenar un clasificador para cada usuario.		Pocas calificaciones por usuario.
Híbridos	Combinar lo mejor de varios mundos.		

Cold Start

New items have no ratings,
new users have no history.

SR • Basado en contenidos



SR • Basado en contenidos

Idea: recomendar ítems al usuario que sean similares a aquellos que puntuó positivamente antes (o, en su defecto, que consumió). Para ello:

1. Para cada ítem, debemos construir un perfil.
 - a. Casos sencillos: información fácilmente disponible. Películas: director, género, actores, año, etc.
 - b. Casos no-sencillos. Debemos extraer features de los ítems. Noticias: hay que usar la batería de herramientas de NLP (tf-idf, etc.)
2. Idealmente, también hay que construir un perfil de qué cosas le gustan al usuario.
3. Usamos una métrica de distancia para encontrar ítems similares.
 - a. Índice Jaccard
 - b. Distancia coseno
4. Recomendamos

SR • Basado en contenidos



No necesitamos información de otros usuarios. (Sin *Cold-Start*)

Puede recomendar a usuarios con “paladar exquisito” o único

Puede recomendar ítems nuevos o poco populares (basta ver su contenido)

Explicable



Hay que armar el perfil de los ítems.
Puede ser difícil encontrar buenos features.

Difícil recomendar a nuevos usuarios.

Puede ser muy específico: no recomienda ítems por fuera perfil del usuario. El usuario puede tener muchos intereses.

SR • Filtro colaborativo

	P1	P2	P3	P4	P5	P6	...	P _m
Usuario 1	5	4			2		...	1
Usuario 2	2	1		5			...	5
Usuario 3		1	5		4	3	...	2
Usuario 4	4			2	1			
...	
Usuario n	1	2	5		5		...	3

SR • Filtro colaborativo

	P1	P2	P3	P4	P5	P6	...	P _m
Usuario 1	5	4	?	?	2	?	...	1
Usuario 2	2	1	?	5	?	?	...	5
Usuario 3	?	1	5	?	4	3	...	2
Usuario 4	4	?	?	2	1	?	...	?
...
Usuario n	1	2	5	?	5	?	...	3

SR • Filtro colaborativo



Funciona para cualquier tipo de ítem (películas, libros, música, etc.).

Puede recomendar ítems por fuera del *perfil* del usuario



Necesitamos la matriz de utilidad

La matriz de utilidad está, en general, vacía y es muy grande. Esto trae dificultades computacionales.

No puede recomendar ítems que no hayan sido calificados previamente

Tiende a recomendar ítems populares

¿Cómo funciona un filtro colaborativo?



SR • Filtro colaborativo • Descomposición UV

Reducción de dimensionalidad - Descomposición UV

$$\begin{bmatrix} 5 & 2 & 4 & 4 & 3 \\ 3 & 1 & 2 & 4 & 1 \\ 2 & & 3 & 1 & 4 \\ 2 & 5 & 4 & 3 & 5 \\ 4 & 4 & 5 & 4 & \end{bmatrix} = \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \\ u_{31} & u_{32} \\ u_{41} & u_{42} \\ u_{51} & u_{52} \end{bmatrix} \times \begin{bmatrix} v_{11} & v_{12} & v_{13} & v_{14} & v_{15} \\ v_{21} & v_{22} & v_{23} & v_{24} & v_{25} \end{bmatrix}$$

SR • Filtro colaborativo • Descomposición UV

Reducción de dimensionalidad - Descomposición UV

$$\begin{bmatrix} 5 & 2 & 4 & 4 & 3 \\ 3 & 1 & 2 & 4 & 1 \\ 2 & & 3 & 1 & 4 \\ 2 & 5 & 4 & 3 & 5 \\ 4 & 4 & 5 & 4 & \end{bmatrix} = \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \\ u_{31} & u_{32} \\ u_{41} & u_{42} \\ u_{51} & u_{52} \end{bmatrix} \times \begin{bmatrix} v_{11} & v_{12} & v_{13} & v_{14} & v_{15} \\ v_{21} & v_{22} & v_{23} & v_{24} & v_{25} \end{bmatrix}$$

5x5 U: 5x2 V: 2x5

SR • Filtro colaborativo • Descomposición UV

Reducción de dimensionalidad - Descomposición UV

$$\begin{bmatrix} 5 & 2 & 4 & 4 & 3 \\ 3 & 1 & 2 & 4 & 1 \\ 2 & & 3 & 1 & 4 \\ 2 & 5 & 4 & 3 & 5 \\ 4 & 4 & 5 & 4 & \end{bmatrix} = \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \\ u_{31} & u_{32} \\ u_{41} & u_{42} \\ u_{51} & u_{52} \end{bmatrix} \times \begin{bmatrix} v_{11} & v_{12} & v_{13} & v_{14} & v_{15} \\ v_{21} & v_{22} & v_{23} & v_{24} & v_{25} \end{bmatrix}$$

5x5

U: 5x2

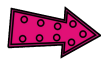
V: 2x5

d: lo elegimos, es un hiperparámetro

SR • Filtro colaborativo • Descomposición UV

Reducción de dimensionalidad - Descomposición UV

$$\begin{bmatrix} 5 & 2 & 4 & 4 & 3 \\ 3 & 1 & 2 & 4 & 1 \\ 2 & & 3 & 1 & 4 \\ 2 & 5 & 4 & 3 & 5 \\ 4 & 4 & 5 & 4 & \end{bmatrix} = \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \\ u_{31} & u_{32} \\ u_{41} & u_{42} \\ u_{51} & u_{52} \end{bmatrix} \times \begin{bmatrix} v_{11} & v_{12} & v_{13} & v_{14} & v_{15} \\ v_{21} & v_{22} & v_{23} & v_{24} & v_{25} \end{bmatrix}$$



Buscamos u_{ij} y v_{ij} de forma que cuando multipliquemos las matrices se aproximen bastante a los valores originales. Ej: $5 = u_{11} * v_{11} + u_{12} * v_{21}$

SR • Filtro colaborativo • Descomposición UV

Reducción de dimensionalidad - Descomposición UV

$$\begin{bmatrix} 5 & 2 & 4 & 4 & 3 \\ 3 & 1 & 2 & 4 & 1 \\ 2 & \bigcirc & 3 & 1 & 4 \\ 2 & 5 & 4 & 3 & 5 \\ 4 & 4 & 5 & 4 & \end{bmatrix} = \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \\ u_{31} & u_{32} \\ u_{41} & u_{42} \\ u_{51} & u_{52} \end{bmatrix} \times \begin{bmatrix} v_{11} & v_{12} & v_{13} & v_{14} & v_{15} \\ v_{21} & v_{22} & v_{23} & v_{24} & v_{25} \end{bmatrix}$$

Buscamos u_{ij} y v_{ij} de forma que cuando multipliquemos las matrices se aproximen bastante a los valores originales. Ej: $5 = u_{11} * v_{11} + u_{12} * v_{21}$



Para completar los lugares vacíos, simplemente ponemos lo que de la multiplicación de la derecha. Ej: $\bigcirc = u_{31} * v_{12} + u_{32} * v_{22}$

SR • Filtro colaborativo • Descomposición UV

Reducción de dimensionalidad - Descomposición UV

¿Cómo encontramos los valores para U y V?

1. Necesitamos una métrica para minimizar. En general, RMSE para los **valores no nulos de la matriz**.
2. Empezamos en algún lugar al azar.
3. Buscamos el mínimo de la función de costo

¡ Es el problema que resuelve el descenso por gradiente !

A close-up photograph of a white ceramic cup filled with a latte. The surface of the milk is decorated with intricate latte art, featuring a central heart shape surrounded by concentric, wavy lines. The cup is placed on a matching white saucer. In the background, a white napkin and a silver spoon are visible, though they are out of focus. The overall lighting is soft and even, highlighting the textures of the coffee and the smooth surface of the cup.

¡BREAK!

Ph. Credit: Drew Coffmann



Hands-on training



Hands-on training

DS_Encuentro_43_Sistemas_Recomendacion.ipynb

Este notebook es la continuación del notebook de la clase anterior. Recomendamos trabajar con ese si aún no lo terminaron.



Recordatorio sobre condiciones de aprobación



Proyectos • Condiciones de aprobación

Los/as evaluadores/as considerarán una entrega como **Aprobada** cuando el/la estudiante haya cumplido satisfactoriamente el 100% de los puntos que pide el checklist (aunque no los haya hecho a todos perfectos).

Caso contrario, el/la evaluador/a considerará la entrega como **Para rehacer.**

Nota: no hay un límite de iteraciones (el/la estudiante puede tener que rehacer su trabajo todas las veces que sea necesario hasta aprobar).

Para la próxima

1. Completar el notebook de hoy.

2. Trabajar en la Entrega 06.

3. Ver el siguiente video:

<https://www.youtube.com/watch?v=kMXZbDT5vm0>

ACÀMICA