

ACÀMICA

TEMA DEL DÍA

Descomposición UV

Una técnica algebraica para predecir los valores faltantes en una matriz de utilidad. Se utiliza para generar sistemas de recomendación con filtros colaborativos.



Agenda

Daily

Explicación: Sistemas de recomendación 2.0

Break

Hands-on training

Proyecto

Cierre



Daily



Daily



Sincronizando...

Bitácora



¿Cómo te ha ido?
¿Obstáculos?
¿Cómo seguimos?

Challenge



¿Cómo te ha ido?
¿Obstáculos?
¿Cómo seguimos?

Repaso de la bitácora

Matriz de utilidad



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 <i>n</i>	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

El objetivo del sistema de recomendación es *poblar* la 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.

Ejemplo: Netflix tiene 150 millones suscriptores, 5 mil películas. La matriz tiene 750.000.000.000 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 750.000.000.000 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 750.000.000.000 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 750.000.000.000 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.

Sistemas de recomendación 2.0



Hoy nos inspiramos con



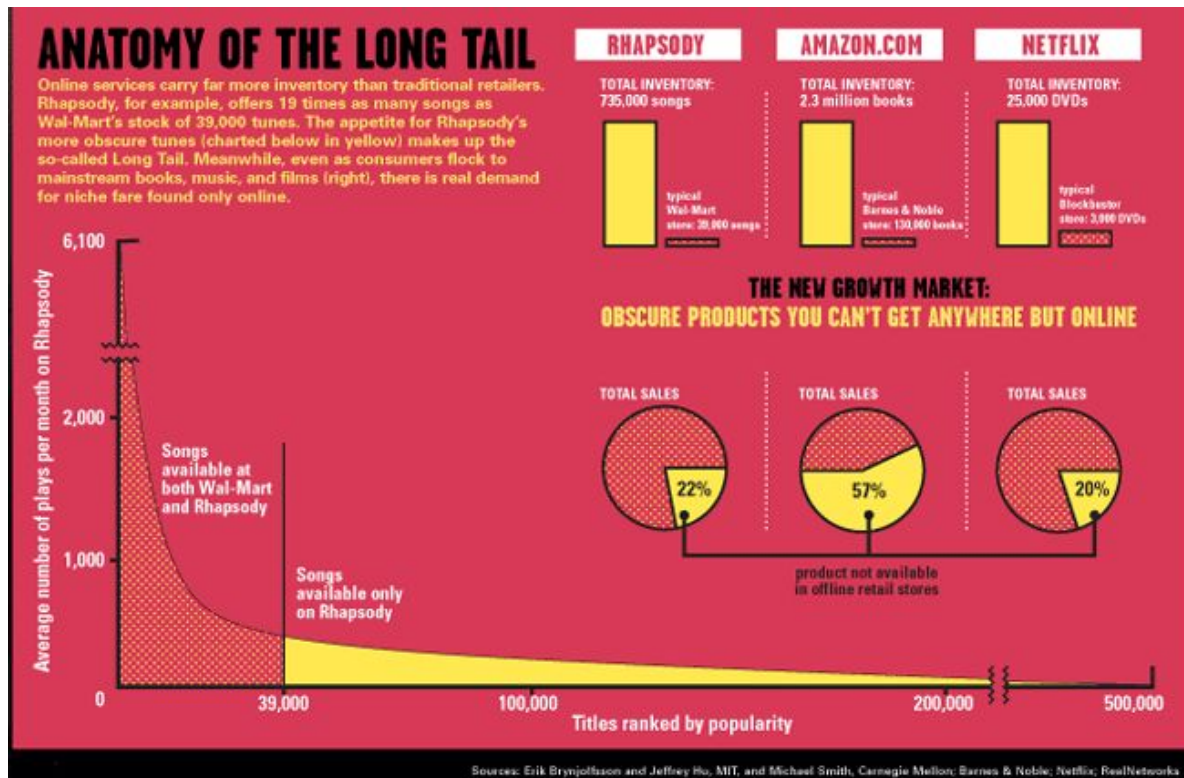
Mining of Massive Datasets

Jure Leskovec, Anand Rajaraman, Jeff Ullman

<http://www.mmds.org/>

The long tail

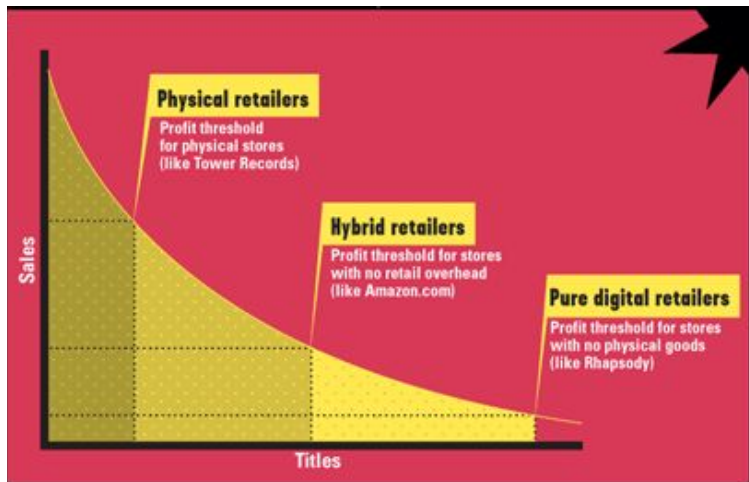
Ejemplo: Into Thin Air y Touching Void



Fuente

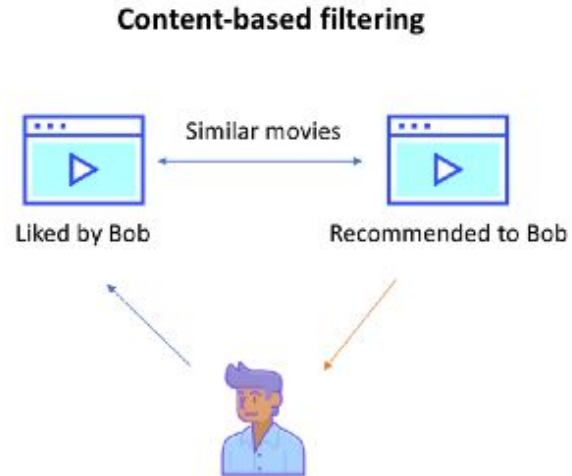
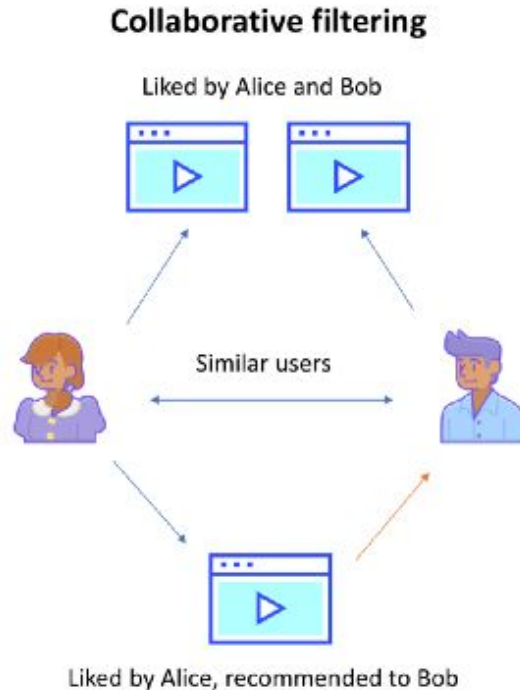
The long tail

Ejemplo: Into Thin Air y Touching Void



Fuente

Tipos de Sistemas de Recomendación



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

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

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

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 de interés para el usuario.
Filtro colaborativo	Recomienda ítems basados en la similitud entre usuarios y/o usuarios e ítems.	matriz de utilidad	1) requiere una gran cantidad de datos de utilidad. 2) puede recomendar ítems que no son de interés para el usuario.
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.

¿Cómo funciona un filtro colaborativo?



Vamos a contarles

Descomposición UV

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

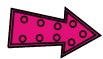
V: 2x5

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?

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**.

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.

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

¿Les suena?



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 !

Evaluación

Train - Test Split

Una buena opción:


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

Evaluación

Train - Test Split

Una buena opción:

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



Test

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_Bitácora_38_SR_SVD.ipynb



Tercer Proyecto



El Proyecto 3 de tu carrera tiene como objetivo que apliques las herramientas aprendidas dentro de tres dominios específicos sobre los que aprenderás en el Bloque 3:

- Sistemas de Recomendación
- Procesamiento de lenguaje natural
- Series de tiempo

El Proyecto 3 de tu carrera tiene como objetivo que apliques las herramientas aprendidas dentro de tres dominios específicos sobre los que aprenderás en el Bloque 3:

- **Sistemas de Recomendación**
- Procesamiento de lenguaje natural
- Series de tiempo

Implementa un Sistema de Recomendación para videojuegos de la plataforma [Steam](#). Apóyate en lo que aprendiste en las bitácoras y durante los encuentros. Además, anímate a investigar y probar más herramientas por tu cuenta.

- Descarga el [dataset aquí](#) (¡Todo el mérito a sus creadores!)
- [Notebook inicial](#)



Para la próxima

- Avanza con el notebook de hoy.
- Lee la bitácora 39 y carga las dudas que tengas al Trello.

En el encuentro que viene uno/a de ustedes será seleccionado/a para mostrar cómo resolvió el challenge de la bitácora. De esta manera, ¡aprendemos todos/as de (y con) todas/as, así que vengan preparados/as.

ACÀMICA