



UNIVERSIDAD
DE GRANADA

Sesión 12: Sistemas de Recomendaciones

Iván Palomares Carrascosa

Carlos Porcel Gallego

25/11/2020



Contenidos

- Introducción.
- Esquemas basados en contenido.
- Tendencias I.
- Esquemas colaborativos.
- Evaluación de sistemas de recomendación.
- Tendencias II.
- Conclusiones.

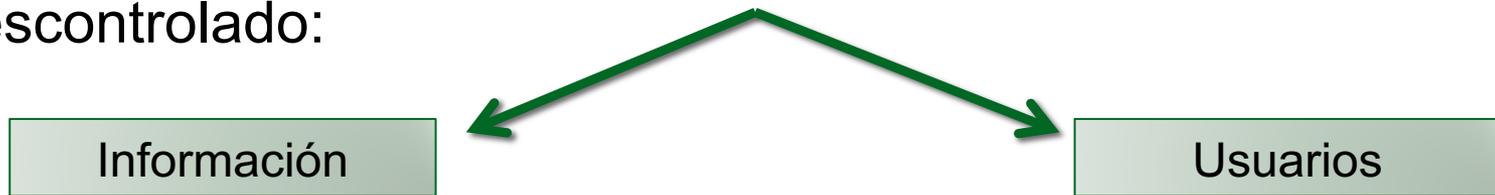
Contenidos

- **Introducción.**
- Esquemas basados en contenido.
- Tendencias I.
- Esquemas colaborativos.
- Evaluación de sistemas de recomendación.
- Tendencias II.
- Conclusiones.

Introducción

Problemas de acceso a la información

- Web: principal fuente de generación y transmisión de la información.
- **Problema** de la Web → crecimiento exponencial y descontrolado:



Introducción

Problemas de acceso a la información

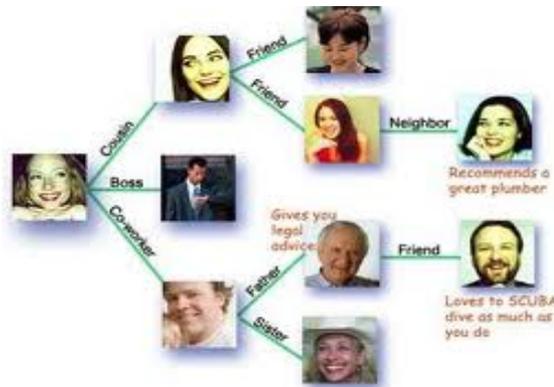
- **Consecuencia:** los usuarios tienen serias dificultades para acceder a la información.
- **Solución:** Desarrollo de sistemas automáticos de búsqueda y acceso a la información en la Web. Dos categorías:
 - Basados en recuperación de información tradicional: consultas de términos.
 - Características de la información: palabras clave.
 - Buscadores Web.
 - Basados en filtrado de información: perfiles de usuarios.
 - Características del usuario: establecimiento de perfiles de usuario.
 - Sistemas de recomendaciones (RecSys).



Introducción

Definición de sistemas de recomendaciones

- **RecSys**: herramientas de filtrado de información que asisten a los usuarios en sus procesos de acceso a la información, mediante la predicción y recomendación de ítems de información que pueden serles de interés.
- Alternativa al proceso social de recomendación: acto habitual que todos practicamos al recurrir a las opiniones de conocidos o expertos cuando tenemos que tomar una decisión para adquirir algo sin tener la suficiente información para ello.



Introducción

- Van más allá de la búsqueda: No responden a necesidades puntuales de información.
- Nos adentran en la Web del descubrimiento:



“los usuarios ya no buscan por algo en concreto, si no que esperan a descubrir cosas que ni siquiera sabían que existían, o no sabían cómo realizar la consulta para encontrar”

- **Serendipia**: encontrar cosas que nos interesan cuando estamos buscando algo relacionado.



Introducción

- Van más allá de la búsqueda: No responden a necesidades puntuales de información.
- Nos adentran en la Web del descubrimiento:



“los usuarios ya no buscan por algo en concreto, si no que esperan a descubrir cosas que ni siquiera sabían que existían, o no sabían cómo realizar la consulta para encontrar”

- **Serendipia**: encontrar cosas que nos interesan cuando estamos buscando algo relacionado.



ME GUSTA

Introducción

- Concepto clave: conseguir una efectiva **personalización**.



- Los RecSys serán clave en el desarrollo de la Web: uso conjunto de RecSys y social media.

Recomendar teniendo en cuenta las opiniones vertidas en redes sociales: **recomendaciones sociales.**



Introducción

RecSys y el comercio electrónico

- Objetivos:



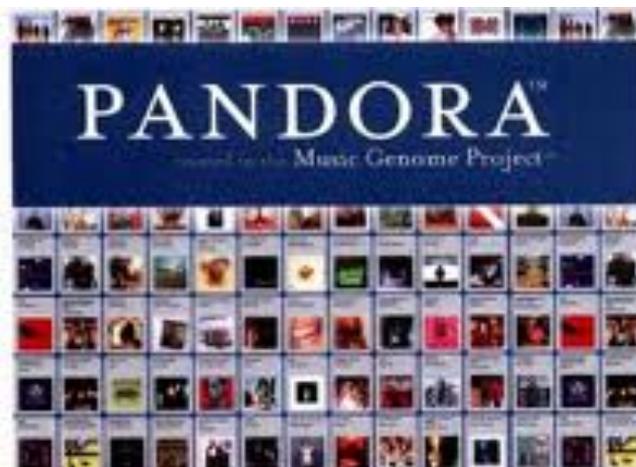
- Ventajas:



Amazon y Netflix generan entre el 30 y 70% de sus ventas a través de las listas de recomendación.

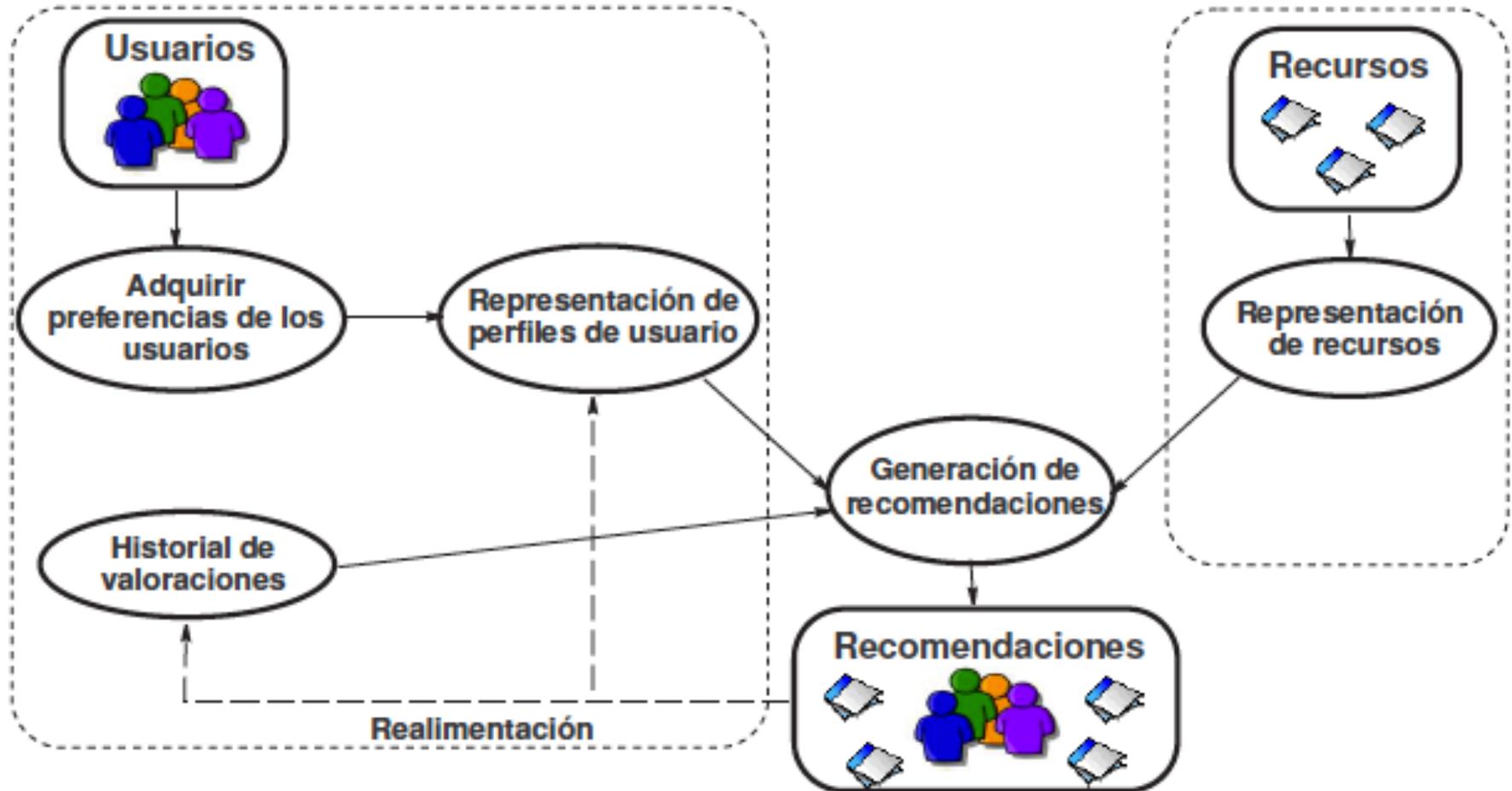
Introducción

Ejemplos



Introducción

Visión general



Introducción

Componentes

Entradas	Salidas
<ul style="list-style-type: none">• Perfil del usuario activo.• Perfiles de los colaboradores.• Representación de los ítems.• Valoraciones dadas a los ítems (feedback).	<ul style="list-style-type: none">• Sugerencia o lista de sugerencias.• Valor de predicción del grado de satisfacción.• Visualizar valoraciones individuales.
Método de generación de recomendaciones	
<ul style="list-style-type: none">• Basados en contenido: al usuario se le recomiendan ítems similares a los que le gustaron en el pasado.• Colaborativos: al usuario se le recomiendan ítems que gustaron a personas con gustos similares.• Híbridos.	

Introducción

Funcionamiento

- Métodos de **estimación de valoraciones** desconocidas, a partir de valoraciones conocidas:
 - **Basados en modelos**: construyen primero un modelo mediante técnicas de aprendizaje automático y luego se consulta el modelo por predicciones de utilidad sobre los ítems.
 - Ejemplo de técnicas usadas: clustering, redes neuronales, árboles de decisión, etc.
 - **Basados en heurísticas o memoria**: sistemas que proporcionan recomendaciones usando fórmulas heurísticas de similitud y correlación entre usuarios e ítems.
 - Ejemplo: similitud basada en coseno.

Introducción

- **Enfoque basado en contenidos**
 - Si al usuario U le gusta el ítem A, y este ítem es similar al ítem B, entonces al usuario U es probable que le guste el ítem B:
 1. **Analizar contenidos**: representación según características.
 2. El **perfil del usuario** activo es representado en función de los ítems que previamente dicho usuario ha probado y valorado.
 3. La generación de recomendaciones se obtiene usando las **características de la representación de los ítems** y un proceso posterior de **cálculo de similitud + agregación**.

Introducción

- Enfoque basado en contenidos: limitaciones
 - Dificultades en analizar de forma automática los contenidos y extracción de características.
 - Sobre especialización: recomendación de contenidos excesivamente relacionados con el perfil de usuario.
 - No diversidad: recomendación de contenidos muy similares.
 - *Cold-start* con nuevos usuarios: un nuevo usuario o no tiene perfil o tiene uno muy poco definido → dificultad para generar recomendaciones para él.

Introducción

- Enfoque colaborativo

- Si al usuario U le gusta el ítem A, y al usuario V le gustan los ítems A y B, entonces puede que al usuario U también le guste el ítem B.
- Perfiles de usuario e ítems son representados mediante patrones de valoraciones numéricas → **Matriz de valoraciones:**

	Ítem 1	...	Ítem n
Usuario 1	-	...	4
...
Usuario m	1	...	5

Introducción

- Enfoque colaborativo

1. **Establecimiento de perfiles**: basados en las valoraciones proporcionadas.
2. **Identificar grupos de usuarios o ítems afines**: se mide el grado de similitud entre usuarios o ítems atendiendo a las valoraciones suministradas → clasificación kNN – *k Nearest Neighbors*:
 - Colaborativo **basado en usuarios**: considera los usuarios con patrones de evaluación similares.
 - Colaborativo **basado en ítems**: considera los ítems con patrones de evaluación similares.
3. La **generación de recomendaciones** se obtiene a partir de los perfiles y de los grupos identificados + agregación.

Introducción

- Enfoque colaborativo

- Colaborativo **basado en usuarios**: considera los usuarios con patrones de valoración similares.

	Ítem1	Ítem2	ítem3	ítem4	ítem5	ítem6
Usuario activo	4	2	3	3	5	?
Usuario1	5	2	2	4	5	3
Usuario2	1	4	2	4	2	2
Usuario3	2	5	5	5	2	4
Usuario4	5	2	4	3	4	4

Introducción

- Enfoque colaborativo

- Colaborativo **basado en ítems**: considera ítems con patrones de valoración similares.

	Ítem1	Ítem2	ítem3	ítem4	ítem5	ítem6
Usuario activo	4	2	3	3	5	?
Usuario1	5	2	2	4	5	3
Usuario2	1	4	2	4	2	2
Usuario3	2	5	5	5	2	4
Usuario4	5	2	4	3	4	4

Introducción

- Enfoque colaborativo: limitaciones
 - Matriz dispersa: cuando el número de ítems valorados es mucho menor que el número total de ítems.
 - *Cold-start* con nuevos usuarios: un nuevo usuario o no tiene, o tiene muy pocas valoraciones.
 - *Cold-start* con nuevos ítems: sobre un nuevo ítem, aún no hay valoraciones.
 - *Grey sheep*: dificultad de generar recomendaciones para un usuario que tenga gustos poco usuales.

Introducción

- Enfoque híbridos
 - Combinan distintas estrategias para explotar las ventajas y reducir los inconvenientes que cada una de ellas presenta por separado.
 - Permiten suavizar o eliminar el problema de cold-start.
 - Son los más habituales.

Contenidos

- Introducción.
- **Esquemas basados en contenido.**
- Tendencias I.
- Esquemas colaborativos.
- Evaluación de sistemas de recomendación.
- Tendencias II.
- Conclusiones.

Esquemas basados en contenido

- Se usan los atributos descriptivos de los ítems para generar recomendaciones.
- Combinan dos fuentes de datos:
 - Descripción de los ítems en términos de atributos sobre su contenido.
 - Perfil de los usuarios, obtenido a partir del feedback.
- Son **apropiados para aliviar el problema de arranque en frío para nuevos ítems.**
- Al no usar perfil de otros usuarios, se **reduce la diversidad y la novedad**: poco sorprendentes para el usuario.

Esquemas basados en contenido

- Especialmente usados en escenarios donde se dispone de bastante información sobre los atributos.
- En muchos casos se trata de **palabras clave** extraídas de la descripción de los productos.
- Por ello, aunque no exclusivamente, se aplican principalmente en entornos centrados en texto, p.e. recomendaciones de noticias.
- En general, métodos de clasificación de texto y modelado de regresión siguen siendo los más utilizados en estos sistemas.

Esquemas basados en contenido

Componentes básicos

- Pre procesamiento y extracción de **características**:
 - Extracción de características de diversas fuentes para obtener una representación basada en palabras clave: modelo vectorial.
 - Altamente dependiente del dominio.
- Establecimiento de **perfiles de usuario**:
 - Construir un modelo para predecir sus intereses → historial de compra y valoraciones.
 - Feedback explícito (valoraciones) o implícito (actividad de los usuarios).
 - Modelo depende de si el feedback es categórico (p.e. comprar o no, seleccionar o no) o numérico (valoraciones o frecuencia de compra).
- Generación de **recomendaciones**.

Esquemas basados en contenido

Pre procesamiento y extracción de características

- Elección del tipo de **representación**, p.e. multidimensional o bolsa de palabras clave.
- Posibilidad de **ponderación de características**:
 - Considerándolos o no, según su relevancia.
 - Asignar pesos según importancia.
- En casos de información no estructurada, se puede obtener una bolsa de palabras → necesidad de **limpieza**:
 - Eliminar “palabras vacías”: artículos, pronombres, etc.
 - *Stemming*: consolidar variaciones de la misma palabra.
 - Extracción de frases: detectar palabras que suelen aparecer juntas de forma frecuente.

Esquemas basados en contenido

Pre procesamiento y extracción de características

- Tanto la selección como la ponderación, también se podrían realizar de forma supervisada → se tienen en cuenta la valoraciones de los usuarios.
- Una vez hecho esto, en la representación vectorial, los documentos se representan como bolsas de palabras junto con su **frecuencia de aparición** → las palabras que más aparecen son penalizadas porque son menos discriminativas.
- Se realiza de forma **offline**.

Esquemas basados en contenido

Recopilar gustos de los usuarios

- Recopilar lo que le gusta y lo que no a los usuarios.
- Posibles formatos:
 - Valoraciones: se especifican las preferencias. Pueden ser binarias, basadas en intervalos u ordinales.
 - Feedback implícito: seguir actividad de los usuarios.
 - Opiniones de texto: extraer valoraciones implícitas a partir de las opiniones → opinion mining y análisis de sentimientos.
 - Casos: los usuarios especifican ejemplos de ítems en los que están interesados, que se usan como feedback implícito en clasificadores como p.e. vecinos más cercanos.
- Se realiza de forma online, conforme los usuarios interactúan con el sistema.

Esquemas basados en contenido

Perfiles de usuario y filtrado

- Relacionado con el problema de clasificación y de modelado de regresión → modelo de entrenamiento.
- El conjunto de entrenamiento corresponde con descripciones de ítems que se han extraído en la fase previa + valoraciones del usuario sobre esos ítems:
 - Construcción de un modelo de entrenamiento específico para cada usuario → **perfil de usuario**.
- El conjunto de test corresponde con descripciones de ítems que aún hayan sido recomendados o valorados por el usuario.

Esquemas basados en contenido

Perfiles de usuario y filtrado

- El modelo de entrenamiento se usa para generar recomendaciones: predicción de valor o lista ordenada.
- Posibles técnicas:
 - Clasificación según el vecino más cercano.
 - Clasificación Bayesiana.
 - Clasificación basada en reglas.
 - Modelos basados en regresión.
 - Etc.

Contenidos

- Introducción.
- Esquemas basados en contenido.
- **Tendencias I.**
- Esquemas colaborativos.
- Evaluación de sistemas de recomendación.
- Tendencias II.
- Conclusiones.

Tendencias I

Esquemas alternativos aplicados en los últimos años:

- **Social RecSys**: basados en redes sociales.
- **Trust RecSys**: basados en confianza.
- **Context RecSys**: basados en contexto.
- RecSys en plataformas de **e-business**.

Tendencias I

Tendencias: Social RecSys

- Usar la gran cantidad de datos que se genera en la Web para realizar recomendaciones sociales.
- Hacer minería Web de las opiniones dadas en redes sociales: podemos encontrar multitud de referencias a marcas y productos, tanto positivas como negativas.
- Analizar sentimientos: detectar estados de ánimo, gente apasionada o decepcionada, etc.
- Resuelve el problema *cold-start*.

Tendencias I

Tendencias: Trust RecSys

- La idea es realizar recomendaciones teniendo en cuenta los usuarios en los que más confiamos, no de los que sean similares en gustos.
- Es necesario calcular o predecir la confianza que unos usuarios tendrán sobre otros que no conocen, a partir de la confianza depositada en otros sí conocidos.



Tendencias I

Tendencias: RecSys basados en contexto

- Hay dominios en los que considerar sólo los ítems y usuarios no es suficiente.
- Contexto engloba ubicación, tiempo, momento (trabajo, casa, etc.), estados de ánimo, entorno social, gustos recientes, etc.
- La idea es usar esa información para realizar recomendaciones según el contexto concreto:
 - Si ahora estoy en Granada, no me recomiendes restaurantes de Segovia.
 - Si últimamente juego mucho con la Wii, no me recomiendes juegos de PC.



Tendencias I

Tendencias: RecSys en comercio electrónico

- La idea es que por las características de los RecSys, podemos obtener muchos beneficios al introducirlos en las plataformas de e-business.
- Incrementar las ventas de la plataforma.
- Ayudan a mantener clientes fieles.
- Ayudar a la gente a conectar y compartir sus experiencias:
 - Groupon presenta ofertas diarias de minoristas y restaurantes de todo el mundo.



Tendencias I

Sistemas desarrollados: SIRE2IN

- RecSys difuso para ayudar al personal de la OTRI en la gestión y difusión de recursos de investigación.
- Inconvenientes:
 - Sistema basado en cálculo de similitud entre recursos y perfiles de investigadores.
 - Falta de serendipia.
- SIRE2IN mejorado:
 - Enfoque híbrido alternando entre colaborativo y basado en contenidos.
 - Mayor grado de personalización.

Tendencias I

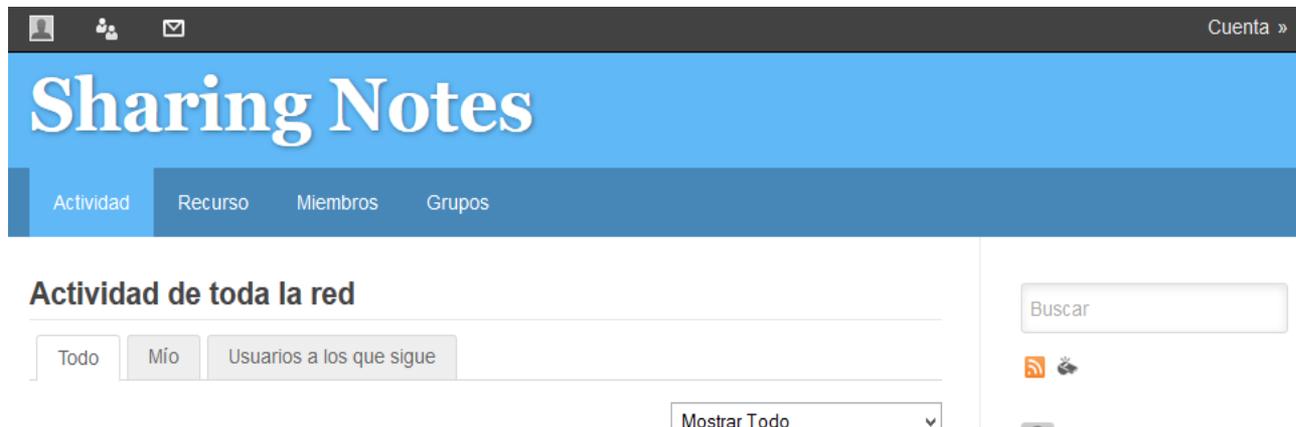
Sistemas desarrollados: Difusión selectiva de información en bibliotecas digitales universitarias

- Cuatro propuestas:
 1. RecSys lingüístico difuso que recomienda tanto recursos especializados como complementarios.
 2. Nuevo método para obtener perfiles de los usuarios → relaciones de preferencia lingüísticas difusas incompletas.
 3. Uso de una memoria para recordar recursos previamente seleccionados pero no recomendados.
 4. Considerar la generación de recomendaciones desde dos puntos de vista: relevancia y calidad.

Tendencias I

Sistemas desarrollados: Social RecSys

- **SharingNotes**: red social docente.
 - Ontologías para representar grados de confianza entre usuarios y las relaciones entre usuarios y sus preferencias.
 - Necesario método para estimar grados de confianza entre usuarios no conectados.
 - Recomendaciones basadas en confianza.



Tendencias I

Sistemas desarrollados: ayuda en tratamiento y prevención de lumbalgias

- **TPLUFIB-WEB** → sistema Web para usuarios con lumbalgia:
 - Ejercicios personalizados y accesibles en cualquier lugar y hora.
 - Basado en modelado lingüístico difuso.
 - Información de confianza: avalada por un equipo de fisioterapeutas y sellos de diversas organizaciones públicas.



Tendencias I

Nuevos desarrollos

- Avanzar en la aplicación de estas técnicas para mejorar la calidad de vida de personas mayores a través del ejercicio físico.
- Mejorar el establecimiento automático de perfiles de usuarios para reducir el impacto del problema de arranque en frío:
 - Extracción de información de redes sociales (Twitter).
 - Análisis de opiniones.

Contenidos

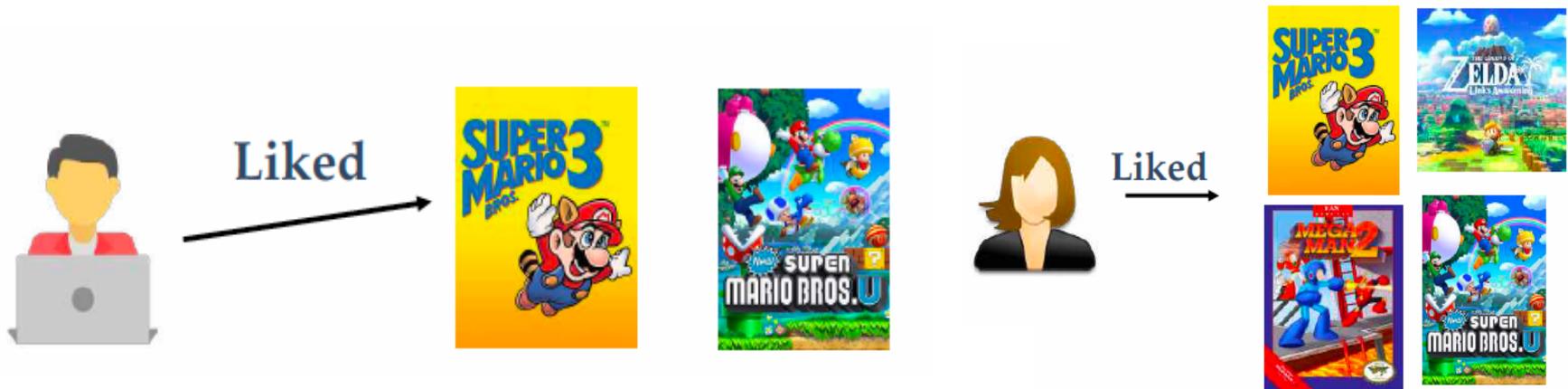
- Introducción.
- Esquemas basados en contenido.
- Tendencias I.
- **Esquemas colaborativos.**
- Evaluación de sistemas de recomendación.
- Tendencias II.
- Conclusiones.

Esquemas colaborativos

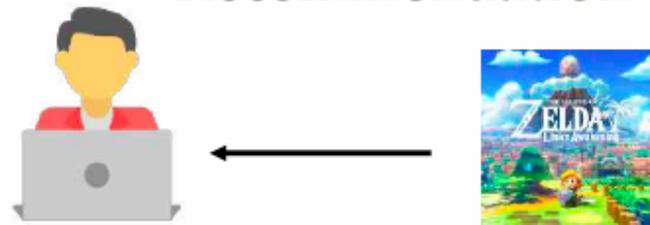
- Se observan las interacciones realizadas por usuarios para hacer recomendaciones.
- Atienden a patrones de *similitud* entre **preferencias** (*gustos, intereses, necesidades, etc.*) de los usuarios o *items*
- Al no estar “sujetos” a similitud entre características de ítems, son **muy útiles para realizar recomendaciones más dinámicas, variadas y novedosas** para el usuario.
- Si se usan de forma aislada, suelen ser propensos al **arranque frío para nuevos usuarios** (*user cold-start*).

Esquemas colaborativos

- Recomendar *ítems* que aún desconozco y que fueron valorados positivamente por usuarios similares a mí



Recommendation



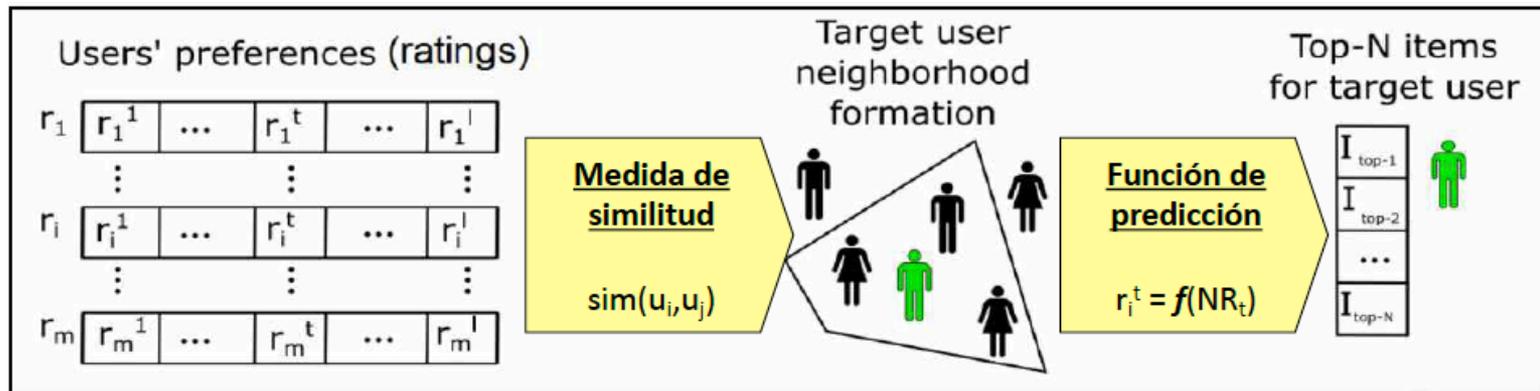
Esquemas colaborativos

- Vecino más cercano (k -NN) basado en usuarios
 - Identifica k usuarios más similares al usuario para el que pretendemos *predecir* su preferencia hacia un ítem.

	Ítem1	Ítem2	ítem3	ítem4	ítem5	ítem6
Usuario activo	4	2	3	3	5	?
Usuario1	5	2	2	4	5	3
Usuario2	1	4	2	4	2	2
Usuario3	2	5	5	5	2	4
Usuario4	5	2	4	3	4	4

Esquemas colaborativos

- Vecino más cercano (k -NN) basado en usuarios
- Identifica k usuarios más similares al usuario para el que pretendemos *predecir* su preferencia hacia un ítem.



Ejemplo: Similitud por coeficiente coseno entre vectores de ratings numéricos

Ejemplo: Índice Jaccard para similitud según ítems marcados como favorito por los usuarios.

Example: Media ponderada de similitudes con k vecinos
NR_t : Valoración dada por vecinos a un ítem "t" desconocido por el usuario "i"

Esquemas colaborativos

- Ejemplo – vecino más cercano (*k*-NN) basado en usuarios
 - Considera los usuarios con patrones de valoración similares.

	Ítem1	Ítem2	ítem3	ítem4	ítem5	ítem6
Usuario activo	4	2	3	3	5	?
Usuario1	5	2	2	4	5	3
Usuario2	1	4	2	4	2	2
Usuario3	2	5	5	5	2	4
Usuario4	5	2	4	3	4	4

$$\mathbf{u} \cdot \mathbf{v} = [u_1 \ u_2 \ \dots \ u_n] \cdot \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = u_1 v_1 + u_2 v_2 + \dots + u_n v_n = \sum_{i=1}^n u_i v_i$$

Esquemas colaborativos

- Ejemplo – vecino más cercano (k -NN) basado en usuarios
 - Considera los usuarios con patrones de valoración similares.

	Ítem1	Ítem2	ítem3	ítem4	ítem5	ítem6
Usuario activo	4	2	3	3	5	?
Usuario1	5	2	2	4	5	3
Usuario2	1	4	2	4	2	2
Usuario3	2	5	5	5	2	4
Usuario4	5	2	4	3	4	4

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\| \|\mathbf{v}\|} = \frac{\sum_{i=1}^n u_i v_i}{\sqrt{\sum_{i=1}^n u_i^2} \sqrt{\sum_{i=1}^n v_i^2}}$$

Esquemas colaborativos

- Ejemplo – vecino más cercano (*k*-NN) basado en usuarios
 - Considera los usuarios con patrones de valoración similares.

	Ítem1	Ítem2	ítem3	ítem4	ítem5	ítem6
Usuario activo	4	2	3	3	5	?
Usuario1	5	2	2	4	5	3
Usuario2	1	4	2	4	2	2
Usuario3	2	5	5	5	2	4
Usuario4	5	2	4	3	4	4

- Similitud entre **usuario activo** y **usuario 1**: 0.981
- Similitud entre **usuario activo** y **usuario 4**: 0.978

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\| \|\mathbf{v}\|} = \frac{\sum_{i=1}^n u_i v_i}{\sqrt{\sum_{i=1}^n u_i^2} \sqrt{\sum_{i=1}^n v_i^2}}$$

Esquemas colaborativos

- Ejemplo – vecino más cercano (*k*-NN) basado en usuarios
 - Considera los usuarios con patrones de valoración similares.

	Ítem1	Ítem2	ítem3	ítem4	ítem5	ítem6
Usuario activo	4	2	3	3	5	?
Usuario1	5	2	2	4	5	3
Usuario2	1	4	2	4	2	2
Usuario3	2	5	5	5	2	4
Usuario4	5	2	4	3	4	4



- Similitud entre **usuario activo** y **usuario 1**: 0.981
- Similitud entre **usuario activo** y **usuario 4**: 0.978
- $\text{Rating}(\text{usuario activo}, \text{item6}) = \frac{0.981*3+0.978*4}{0.981+0.978} = 3.49$

Esquemas colaborativos

- Ejemplo – predicción según carácter (optimismo/pesimismo) del usuario

	Item1	Item2	Item3	Item4	Item5		
Alice	5	3	4	4	?		
User1	3	1	2	3	3		sim = 0,85
User2	4	3	4	3	5		sim = 0,70
User3	3	3	1	5	4		sim = 0,00
User4	1	5	5	2	1		sim = -0,79

$$pred(a, p) = \bar{r}_a + \frac{\sum_{b \in N} sim(a, b) * (r_{b,p} - \bar{r}_b)}{\sum_{b \in N} sim(a, b)}$$

$$4 + 1/(0.85 + 0.7) * (0.85 * (3 - 2.4) + 0.70 * (5 - 3.8)) = 4.87$$

Esquemas colaborativos

Preferencias Explícitas

- El usuario expresa deliberadamente su opinión sobre un ítem conocido/consumido.

- Puntuación (*rating*): 0-10, 5 estrellas, etc.
- *Rating binario*: **me gusta vs no me gusta**
- *Rating unario*: marcadores, favoritos, ...
- Reseñas textuales (*reviews*)



- Los modelos de recomendación más sencillos se basan en *ratings* numéricos.
- Extraer conocimiento de las preferencias a partir de información no numérica –por ejemplo, texto- requiere técnicas adicionales como **Procesamiento de Lenguaje Natural**.

Esquemas colaborativos

Preferencias Implícitas

- No se proporcionan de forma deliberada
- Recolectadas *observando* las interacciones del usuario en el sistema
 - Clicks en páginas del producto o servicio.
 - Compras realizadas.
 - Reproducción de videos o música.
 - Tiempo invertido en un producto concreto.
 - ...



- Los datos de interacción usuario-ítem NO implican necesariamente preferencia positiva hacia el ítem!

Esquemas colaborativos

Factorización de matrices

- Una de las técnicas mas avanzadas de filtrado colaborativo.
- Permite reducir complejidad del problema, obteniendo una representación compacta (modelo) de interacciones *user-ítem*.
- **Reducción de dimensionalidad de los datos.**
- Usado por Youtube, **Netflix** y Yelp.



Esquemas colaborativos

Factorización de matrices

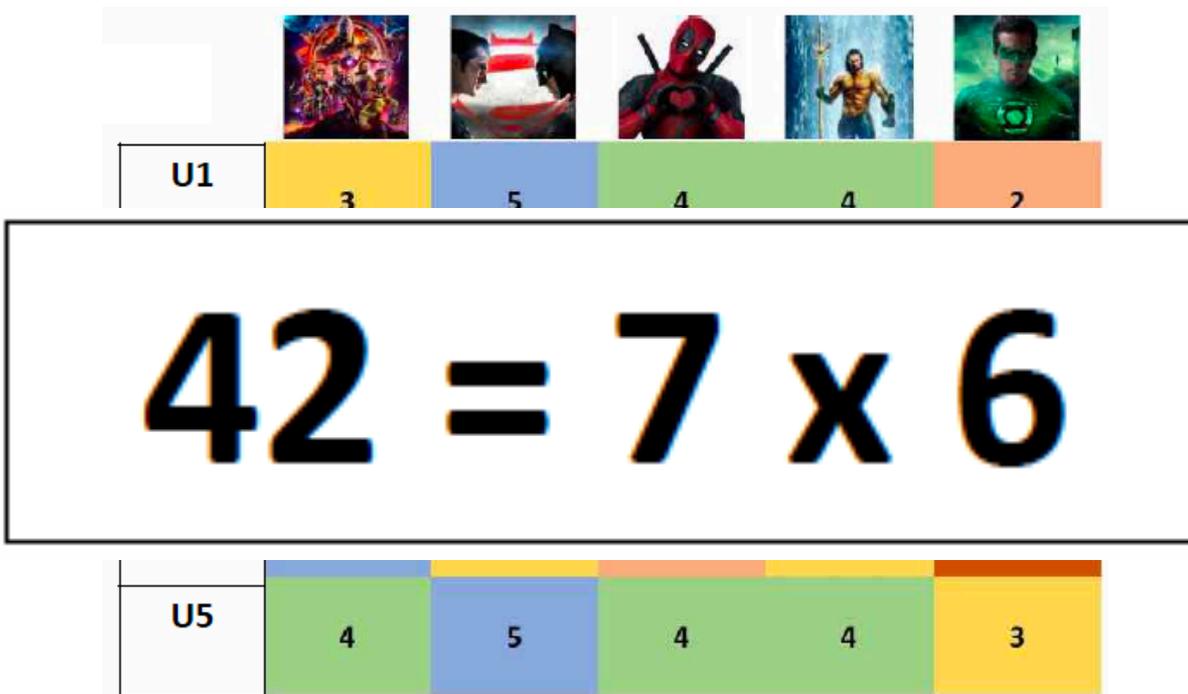
- Intentar predecir si a U4 le gustará *'Green Lantern'*
- **Factorización:** Expresar algo grande (una matriz) como el producto de dos componentes (dos matrices) más pequeños

					
U1	3	5	4	4	2
U2	5	3	4	3	1
U3	3	2	5	3	1
U4	5	3	2	3	???
U5	4	5	4	4	3

Esquemas colaborativos

Factorización de matrices

- Intentar predecir si a U4 le gustará *'Green Lantern'*
- **Factorización:** Expresar algo grande (una matriz) como el producto de dos componentes (dos matrices) más pequeños



Esquemas colaborativos

Factorización de matrices

- Hallar dos matrices rectangulares de **menor dimensión** que la matriz de *ratings* original.
- Estas dos matrices o *factores* conservan la información sobre las propiedades y dependencias entre los **m** usuarios y **n** ítems en la matriz original

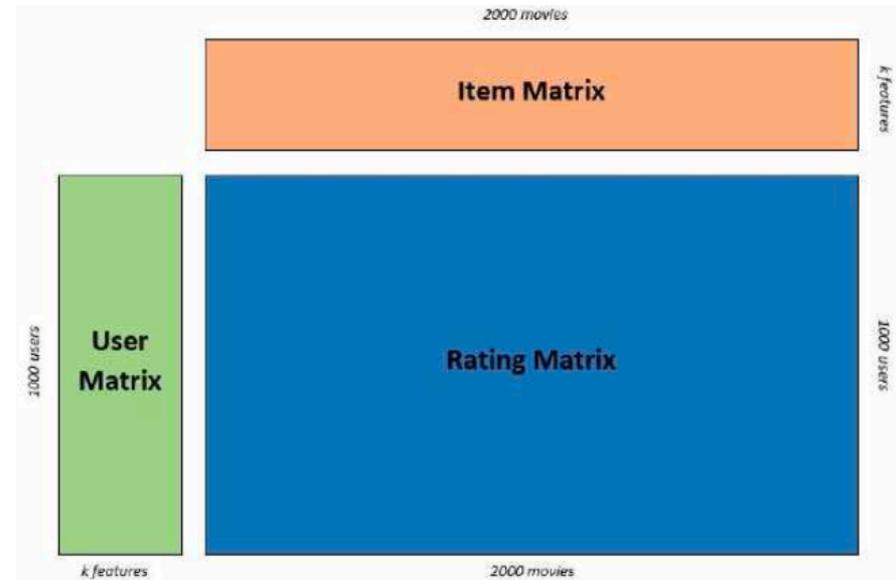
					
U1	3	5	4	4	2
U2	5	3	4	3	1
U3	3	2	5	3	1
U4	5	3	2	3	???
U5	4	5	4	4	3

Esquemas colaborativos

Factorización de matrices

- Hallar dos matrices rectangulares de **menor dimensión** que la matriz de *ratings* original.
 - Estas dos matrices o *factores* conservan la información sobre las propiedades y dependencias entre los **m** usuarios y **n** ítems en la matriz original
- I. Matriz de *usuarios*: m usuarios, k factores latentes.
 - II. Matrix de *ítems*: k factores latentes, n ítems.

					
U1	3	5	4	4	2
U2	5	3	4	3	1
U3	3	2	5	3	1
U4	5	3	2	3	???
U5	4	5	4	4	3

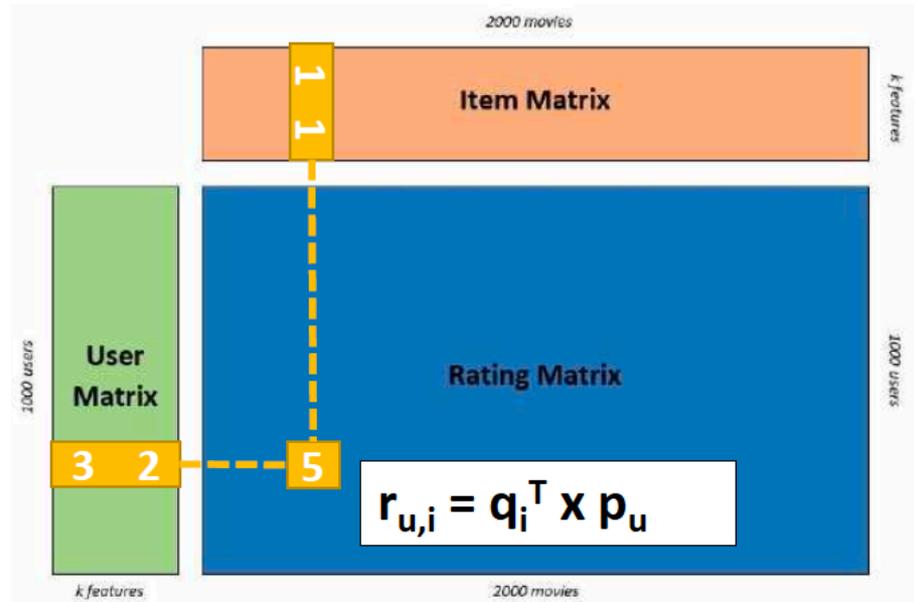


Esquemas colaborativos

Factorización de matrices

- Hallar dos matrices rectangulares de **menor dimensión** que la matriz de *ratings* original.
 - Estas dos matrices o *factores* conservan la información sobre las propiedades y dependencias entre los **m** usuarios y **n** ítems en la matriz original
- I. Matriz de *usuarios*: *m* usuarios, *k* factores latentes.
 - II. Matrix de *ítems*: *k* factores latentes, *n* ítems.

					
U1	3	5	4	4	2
U2	5	3	4	3	1
U3	3	2	5	3	1
U4	5	3	2	3	???
U5	4	5	4	4	3



Esquemas colaborativos

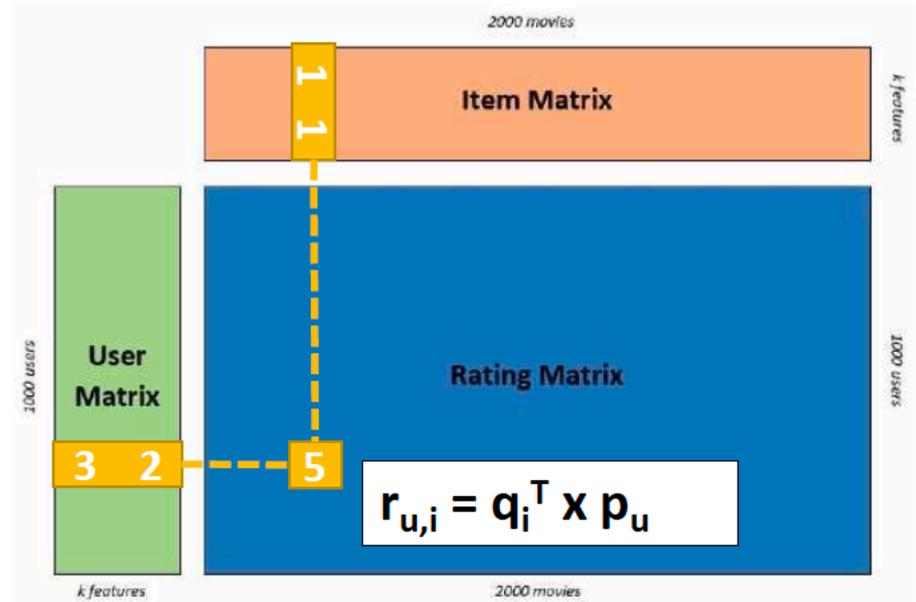
Factorización de matrices

- Hallar dos matrices rectangulares de **menor dimensión** que la matriz de *ratings* original.
- Estas dos matrices o *factores* conservan la información sobre las propiedades y dependencias entre los **m** usuarios y **n** ítems en la matriz original

1. *Rating* en la matriz original = **producto escalar** de ...

- Fila u de matriz de usuarios: p_u
- Columna i de la matriz de ítems: q_i^T

					
U1	3	5	4	4	2
U2	5	3	4	3	1
U3	3	2	5	3	1
U4	5	3	2	3	???
U5	4	5	4	4	3

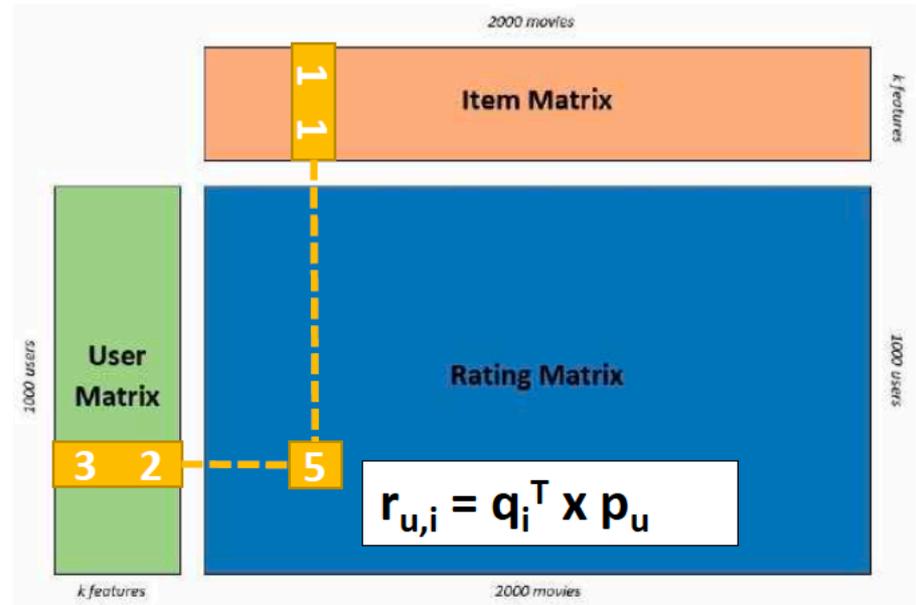


Esquemas colaborativos

Factorización de matrices

- Cómo obtener los vectores de factores latentes → **Optimización**
- Aprender (optimizar) los vectores latentes de las dos matrices que buscamos, de modo que *aproximen* los *ratings* originales de que disponemos:
 - Minimización de error.
 - Técnicas populares:
 - *Gradient Descent, Stochastic Gradient Descent.*
 - *Alternating Least Squares.*

					
U1	3	5	4	4	2
U2	5	3	4	3	1
U3	3	2	5	3	1
U4	5	3	2	3	???
U5	4	5	4	4	3



Esquemas colaborativos

Factorización de matrices

- Los vectores latentes se obtienen para todo usuario e ítem, a pesar de que la matriz de *ratings* original no estará densamente poblada.
 - Esto ayuda, con un **dot product**, a **predecir ratings** y completar la matriz original!

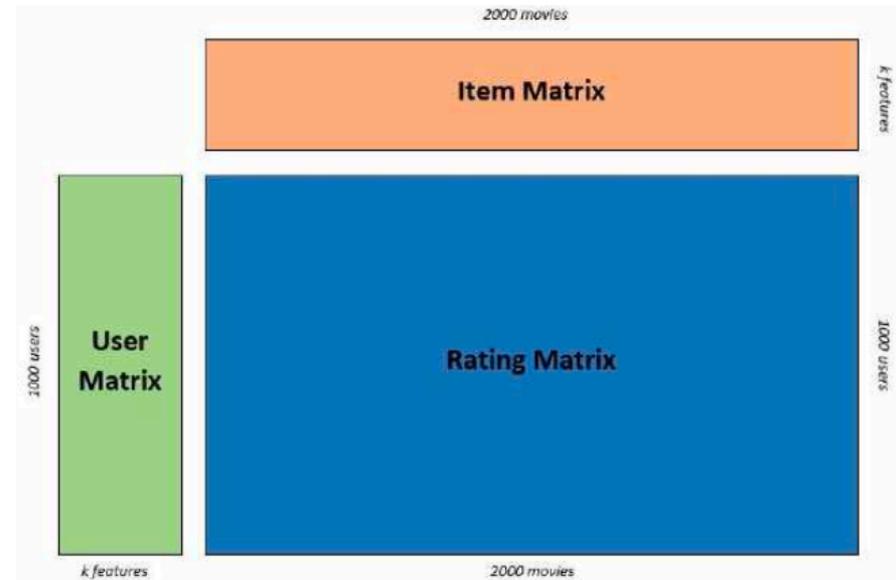
VENTAJAS

- *Alta precisión (accuracy)*
- *Reduce el coste por alta dimensionalidad*

DESVENTAJAS

- *Recalcular modelo periódicamente a medida que los datos reales existentes cambian/aumentan*

					
U1	3	5	4	4	2
U2	5	3	4	3	1
U3	3	2	5	3	1
U4	5	3	2	3	???
U5	4	5	4	4	3



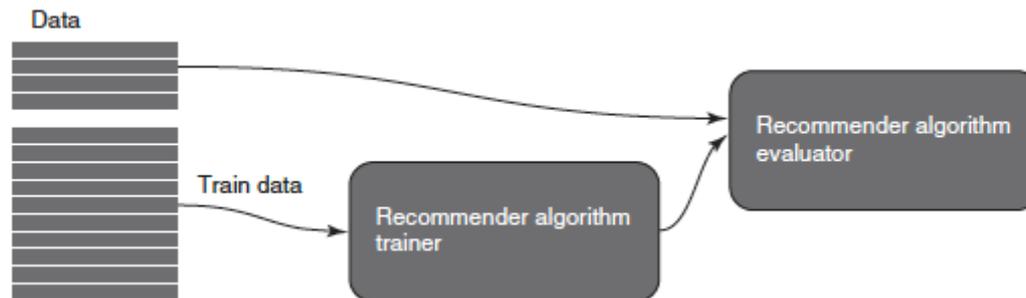
Contenidos

- Introducción.
- Esquemas basados en contenido.
- Tendencias I.
- Esquemas colaborativos.
- Evaluación de sistemas de recomendación.
- Tendencias II.
- Conclusiones.

Evaluación de los sistemas de recomendación

Evaluación Offline

- Comparación (asequible) de diferentes algoritmos de recomendación candidatos → **DATOS**
- No requiere interacción directa con usuarios.
- Menos fiable que estudios reales, sobretodo cuando conviene evaluar la experiencia/percepción del usuario, sesgos...



Fuente: K. Falk. Practical Recommender Systems. Manning, 2019.

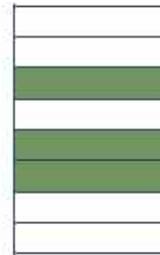
A. Bellogin, A. Said. *Offline and Online Evaluation of Recommender Systems*, Cap. 9, *Collaborative Recommendations: Algorithms, Practical Challenges and Applications*. World Scientific, 2019.

Evaluación de los sistemas de recomendación

Evaluación Offline: Métricas

- **Precisión en N** (N = número de recomendaciones)
- Separar el conjunto de *ratings conocidos* de un usuario en **entrenamiento + test**
- Objetivo ideal: Recomendar ítems –valorados positivamente– en el conjunto *test* del usuario → **éxito**
- “¿Cuántas de las *N* recomendaciones son exitosas?”

Precision at N: $P@N = \frac{\#successful}{N}$



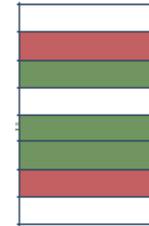
$P@N = 3/8$

Evaluación de los sistemas de recomendación

Evaluación Offline: Métricas

- **Índice de éxito en N** (*success rate at N*)
- Se incluyen tanto predicciones sobre *ratings* ‘de test’ positivos como aquellas sobre *ratings* ‘de test’ negativos.

Success rate at N:
$$S@N = \frac{\#successful}{\#successful + \#unsuccessful} = \frac{3}{3+2} = \frac{3}{5}$$



- **Recall en N**
- “¿Cuántos de los ítems que sabemos que le gustaron al usuario (relevantes) son recomendados?”

$$R@N = \frac{\#successful}{\#known\ successful\ interactions}$$

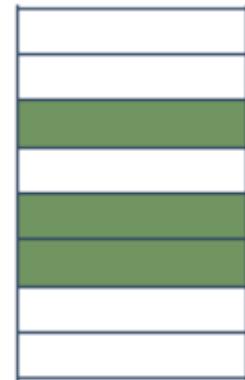
Evaluación de los sistemas de recomendación

Evaluación Offline: Métricas

$$F1 = \frac{2 P \cdot R}{P + R}$$

- **F1 Score**: combina precisión y recall
 - Precisión: Cuántos de los ítems recomendados son relevantes.
 - Recall: Cuántos de los ítems relevantes son recomendados.
- **Métricas de ranking** → **DCG@N** (*Discounted Cumulative Gain*)
- Mide la calidad del ranking realizado, observando las posiciones que ocupan las recomendaciones.

$$DCG@N = \sum_{i=1}^N \frac{2^{rel_i} - 1}{\log_2(i+1)} = \frac{1}{\log_2(3+1)} + \frac{1}{\log_2(5+1)} + \frac{1}{\log_2(6+1)} = 4.12$$



Evaluación de los sistemas de recomendación

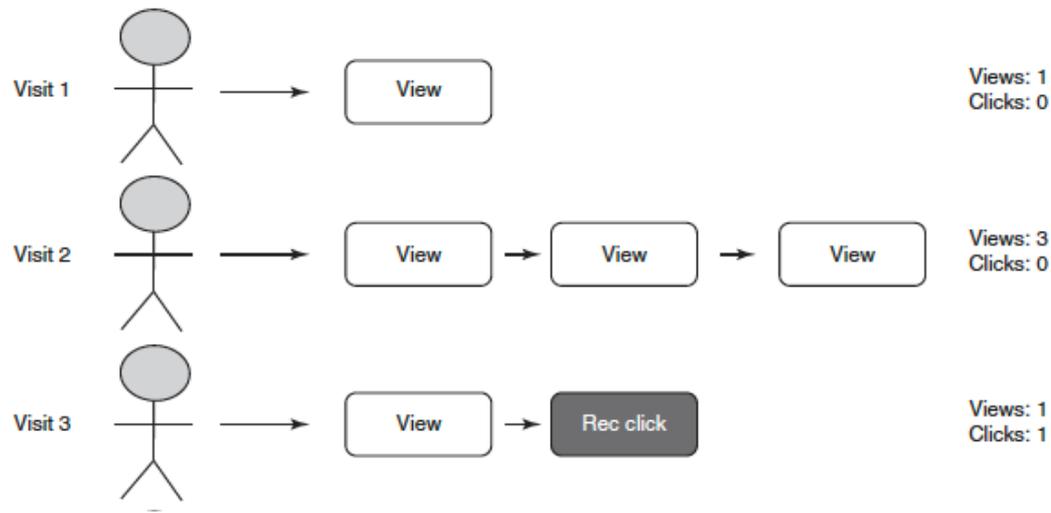
Evaluación Offline: otras propiedades a evaluar

- *Diversidad*: cómo de variadas o repetitivas son las recomendaciones.
- *Serendipia*: cómo de sorprendente o inesperada es la recomendación → novedad + diversidad
- *Coste computacional*: en entornos donde es importante obtener recomendaciones a partir de grandes datos en tiempo real.
- Adecuación al contexto
- ...

Evaluación de los sistemas de recomendación

Evaluación Online

- Participación directa de usuarios → valoración de la calidad de las recomendaciones según la perciben los usuarios
- Evaluar interacciones (*feedback*) al recibir recomendaciones



Fuente: K. Falk. Practical Recommender Systems. Manning, 2019.

Evaluación de los sistemas de recomendación

Evaluación Online

- Enfoque más aceptado → Dos fases

1. Experimentos controlados

- Invitar a un grupo de personas al azar a experimentos en un entorno controlado.
- Cuestionario para recopilar preferencias → envío de recomendaciones → preguntar al usuario (ej. *¿Cuál de estas dos opciones te parece mejor?*)



Marinated Ranch Broiled Chicken
Fat: 15 g / per serve



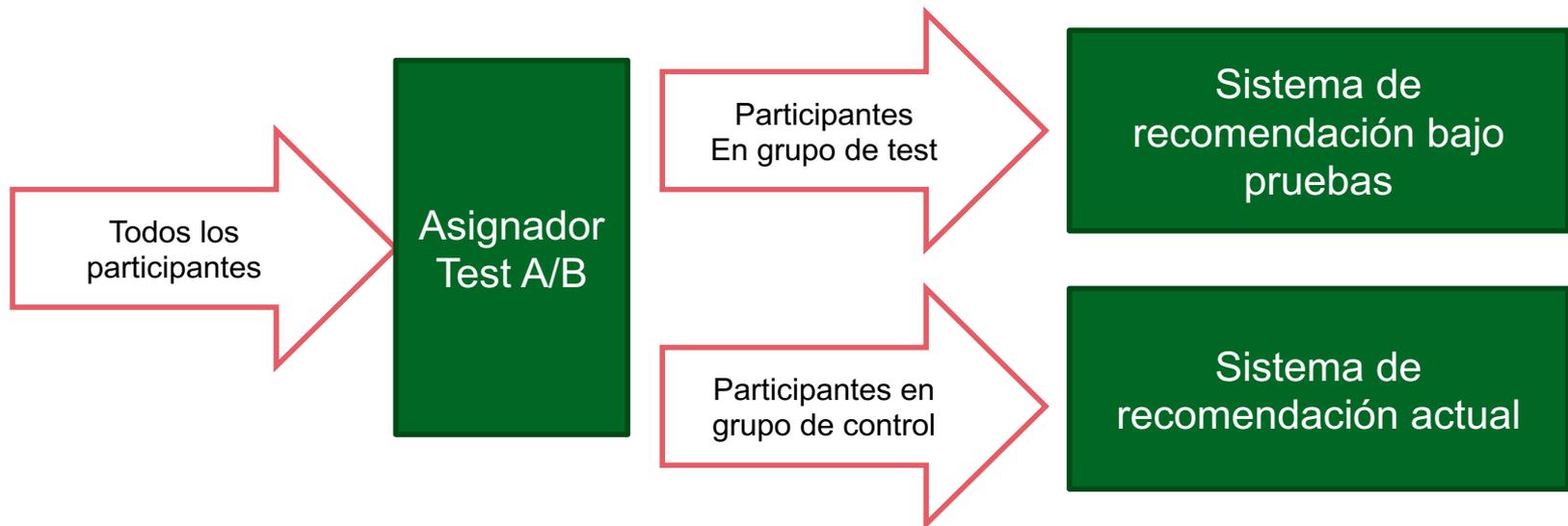
Ranch Crispy Chicken
Fat: 1.7 g / per serve

Créditos imagen: C. Trattner. *Food Recommendations*. Plenary talk, DSRS-Turing 2019 conference.

Evaluación de los sistemas de recomendación

Evaluación Online

- Enfoque más aceptado → Dos fases
 2. Pruebas con usuarios al azar y/o incrementales → Test **A/B**
 - Lanzar dos versiones de un modelo “al azar”, dividir usuarios en dos grupos, y medir cuál funciona mejor (estrategia común en marketing)



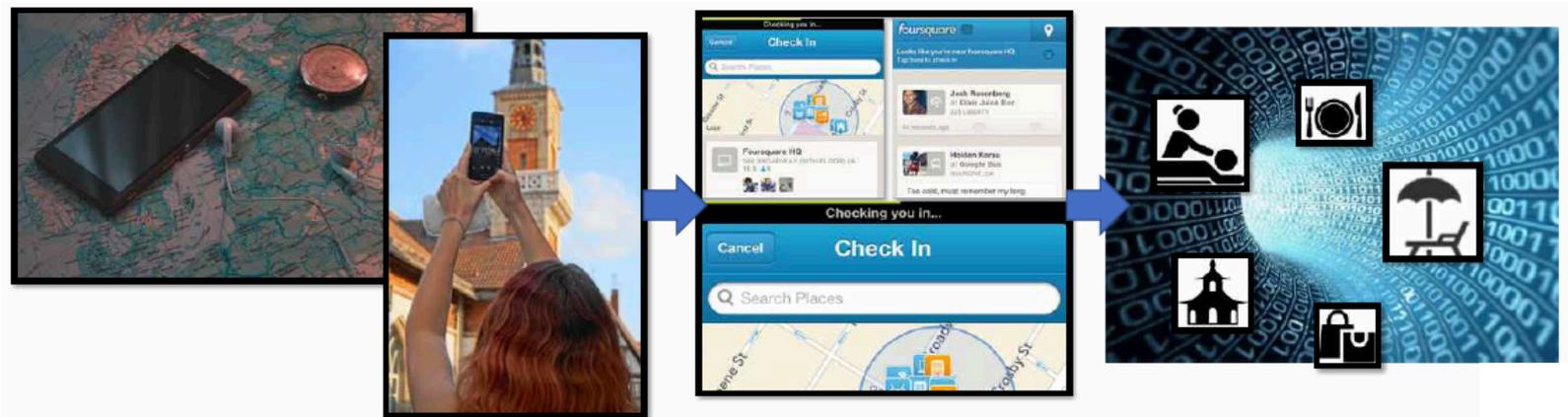
Contenidos

- Introducción.
- Esquemas basados en contenido.
- Tendencias I.
- Esquemas colaborativos.
- Evaluación de sistemas de recomendación.
- Tendencias II.
- Conclusiones.

Tendencias II

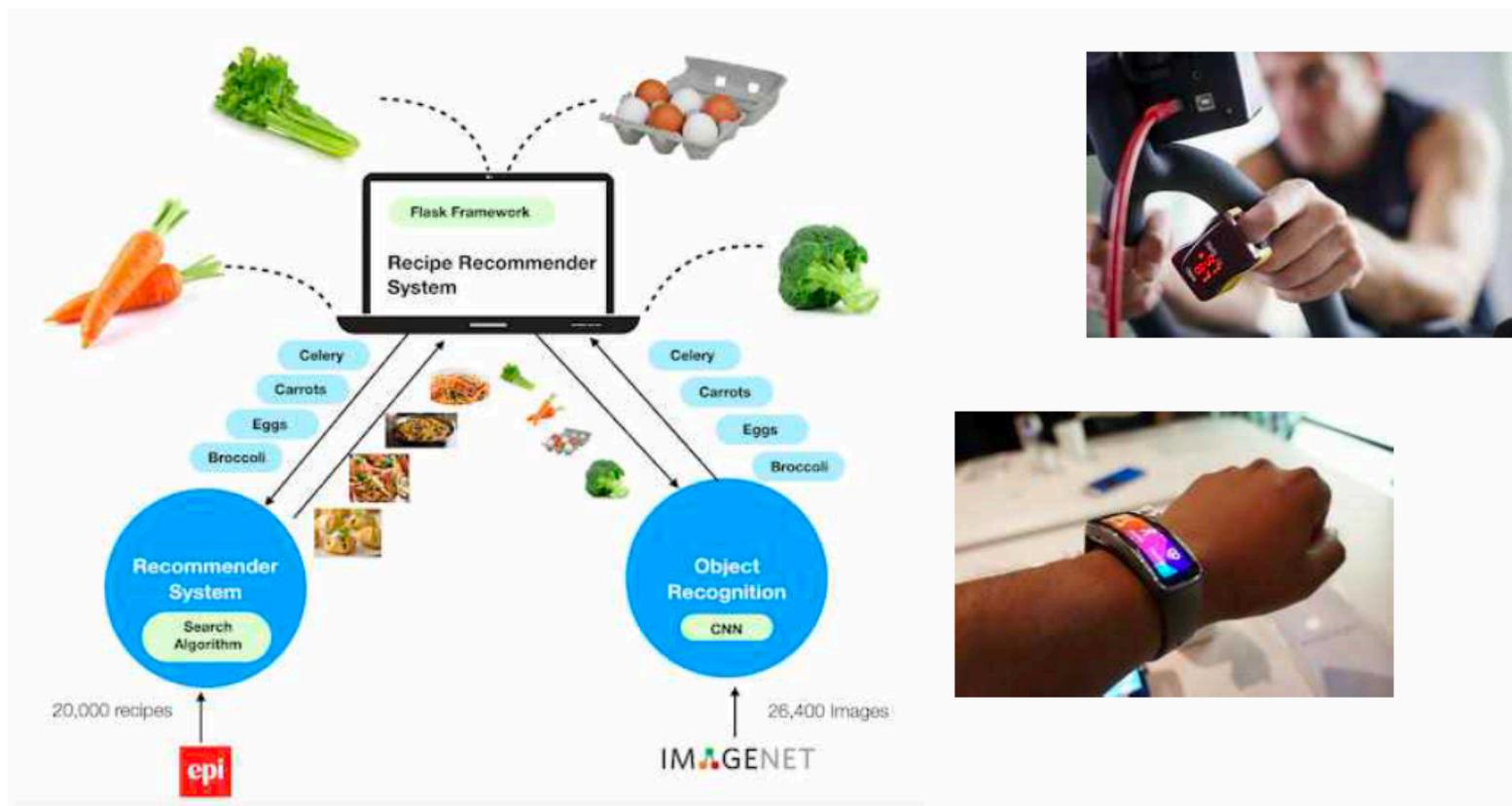
Preferencias y Contexto – Ejemplo geolocalización (turismo)

- Foursquare, Gowalla, Flickr, Facebook Places, ...
- Los usuarios realizan visitas (*check-ins*) o suben fotografías tomadas en *puntos de interés* (PoI) en ciudades y otros enclaves.
- **DATOS:** *Check-in realizados en Pols de cierta(s) categorías, en un contexto determinado.*



Tendencias II

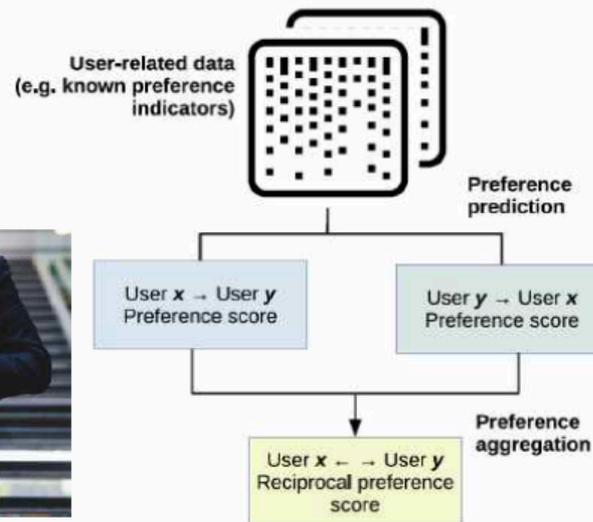
Aplicaciones emergentes – Salud y bienestar



Tendencias II

- **Sistemas de recomendación recíprocos**
 - Recomiendan usuarios a otros usuarios para conectar entre sí.
 - Aparece el requisito añadido de la *reciprocidad* → predicción de **preferencia o compatibilidad mutua**

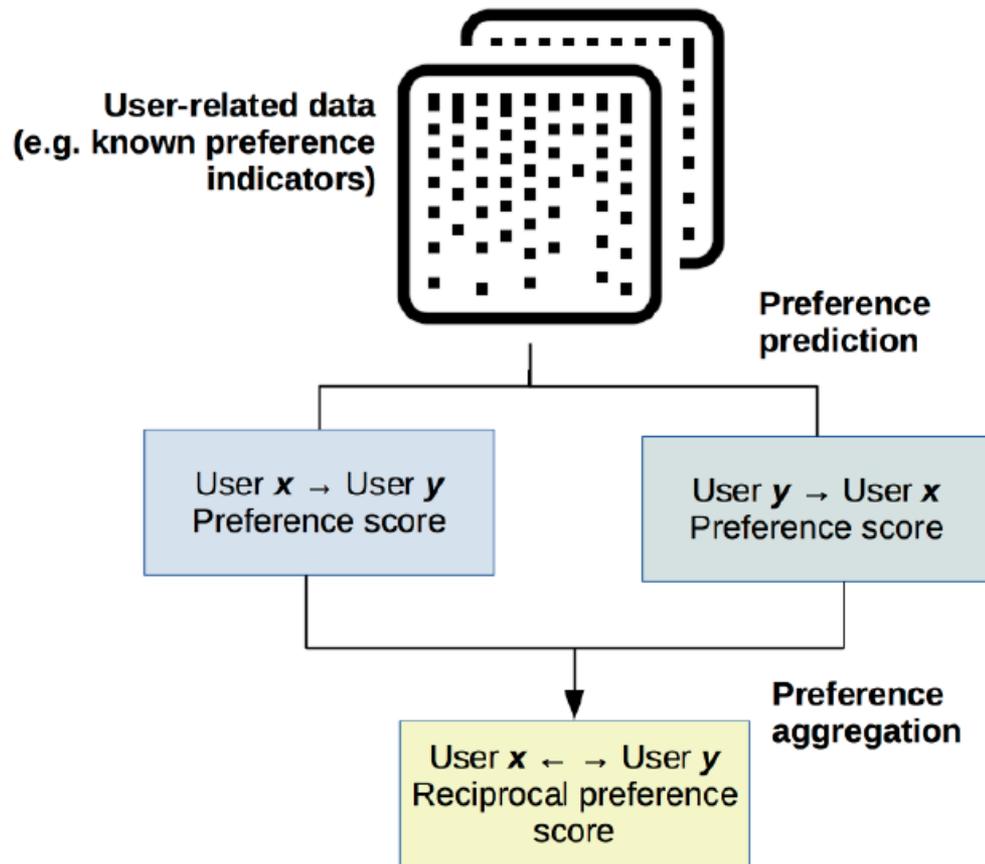
APLICACIONES EMERGENTES – RECOMENDACIONES RECÍPROCAS



I. Palomares, C. Porcel, L. Pizzato, I. Guy, E. Herrera-Viedma. [Reciprocal Recommender Systems: Analysis of State-of-the-Art Literature, Challenges and Opportunities on Social Recommendation](#). ArXiv preprint, Nov 2020.

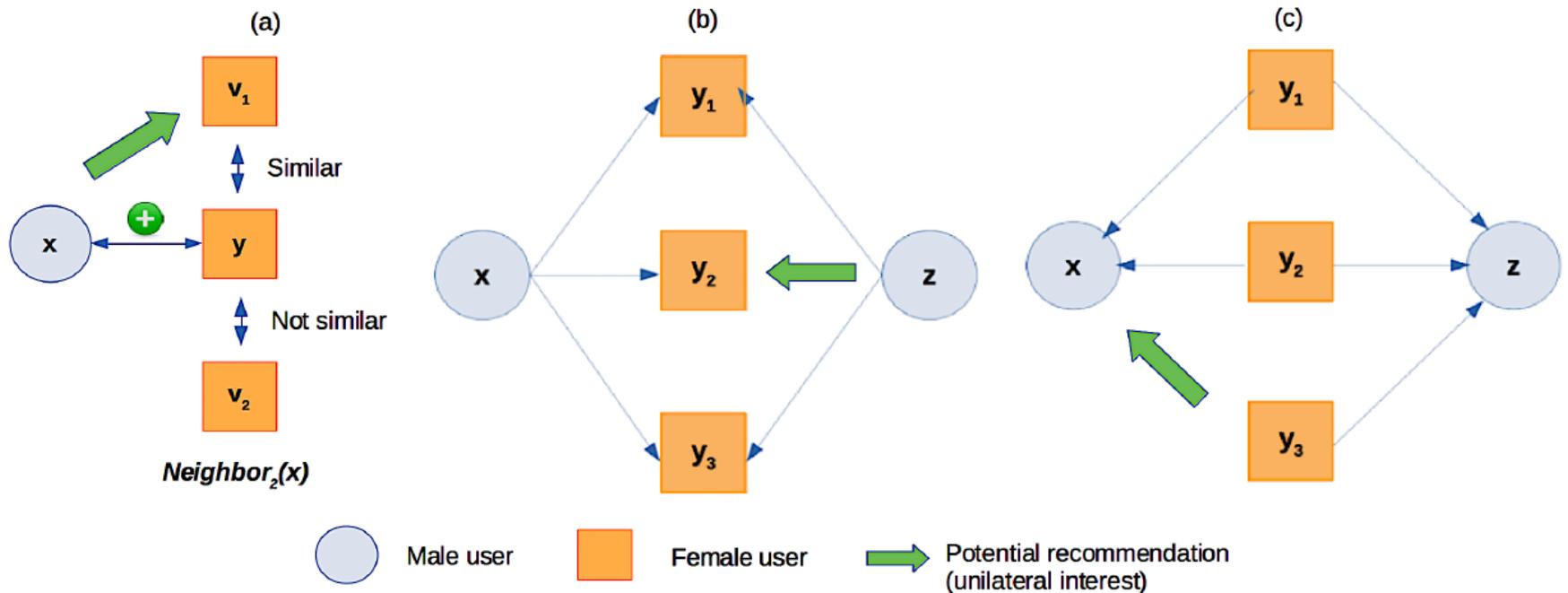
Tendencias II

- Sistemas de recomendación recíprocos



Tendencias II

- Sistemas de recomendación recíprocos
 - a) Esquema basado en contenido
 - b) Esquema colaborativo basado en *similitud de intereses*
 - c) Esquema colaborativo basado en *similitud de atractividad*



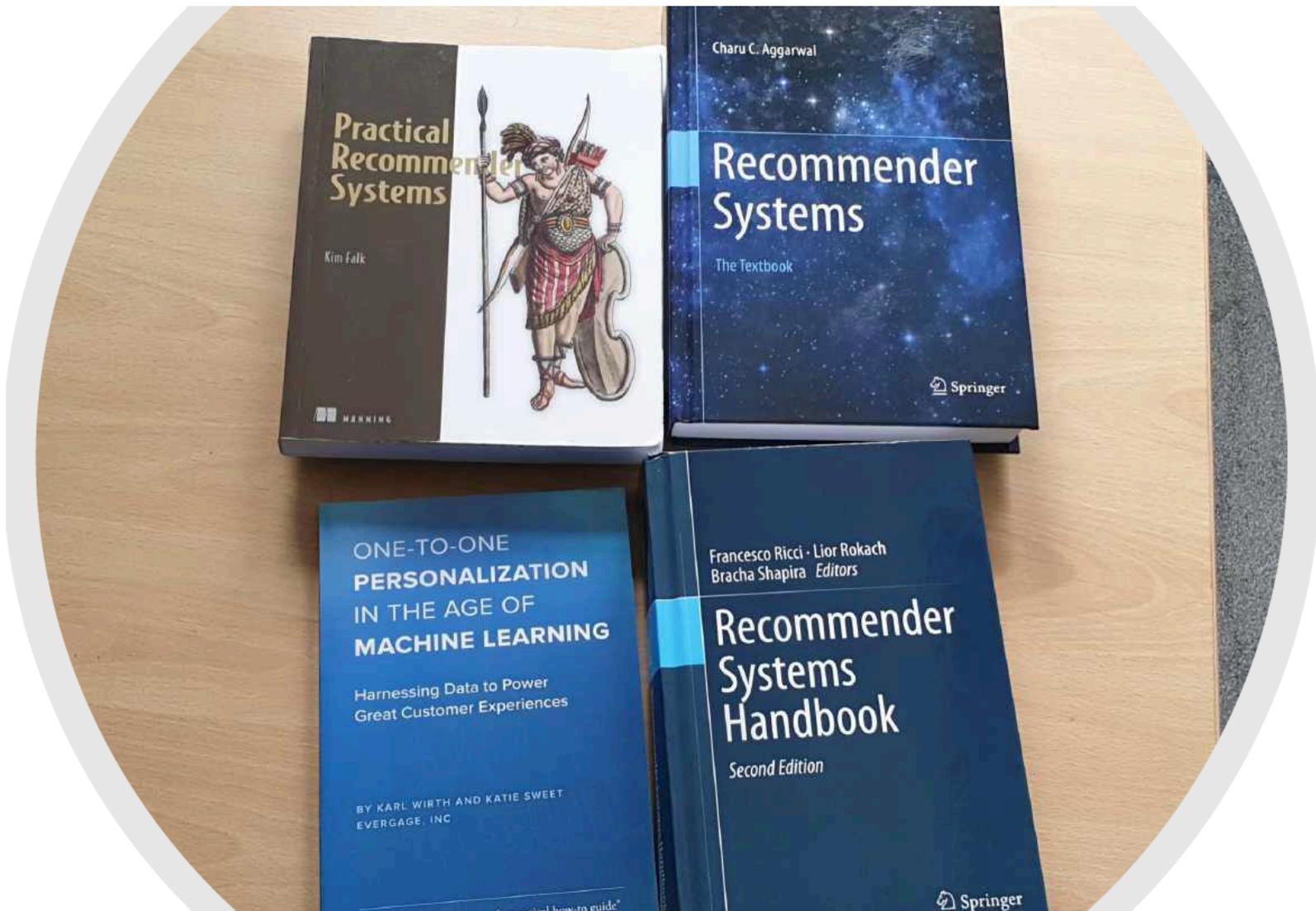
Contenidos

- Introducción.
- Esquemas basados en contenido.
- Tendencias I.
- Esquemas colaborativos.
- Evaluación de sistemas de recomendación.
- Tendencias II.
- Conclusiones.

Conclusiones

- **Sistemas de recomendación**
 - Ayudan al usuario a encontrar información relevante analizando sus gustos, preferencias y necesidades.
 - Rango de aplicaciones cada vez mayor, en línea con el crecimiento de plataformas de comercio, ocio y entretenimiento en Internet, apps turísticas, de salud, de socialización y conexión entre personas, etc.
 - Enfoques basados en contenido, colaborativos, híbridos, basados en contexto, etc. No hay ningún esquema mejor que los demás → depende del ámbito de aplicación concreto.
 - Retos que acompañan a los avances realizados: evaluación, reproducibilidad, ética, enfoques de IA confiable, etc. → necesidad de seguir investigando.

Para saber más...





UNIVERSIDAD
DE GRANADA

Sesión 12: Sistemas de Recomendaciones

Iván Palomares Carrascosa – ivanpc@ugr.es

Carlos Porcel Gallego – cporcel@decsai.ugr.es

25/11/2020

