

Sistemas Recomendadores

IIC-3633

Recomendación no personalizada
Filtrado colaborativo basado en usuario
Filtrado colaborativo basado en ítems

Esta clase

1. User-Based Collaborative Filtering
2. Item-Based Collaborative Filtering

Recomendación no personalizada, semi-personalizada y personalizada.

No
personalizada

Semi- personalizada

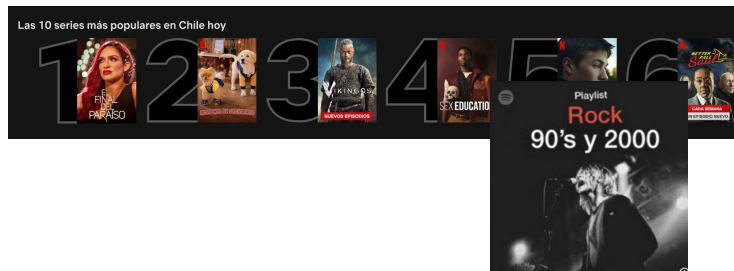
Personalizada



- Más populares
 - Random
 - L2R

Basada en regla
general
(ej País,
Segmentación por
edad, intereses)

Basada en
intereses
personales



- Filtrado colaborativo
- Factores latentes
- Basada en contenido
- Ensamblajes / Híbridos

Recomendación personalizada

Considera información del usuario:

- Ítems con los que ha interactuado históricamente (rating, compra/reproducción, like, etc)

- Información de su perfil

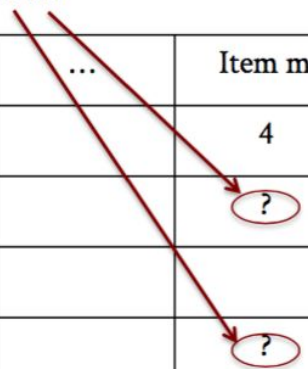
- Información contextual (ubicación , día , hora, etc..)

Predicción de Ratings

- Los algoritmos que veremos el resto de esta clase simplifican el problema de recomendación considerándolo como un problema de predicción de ratings basado en coocurrencia :

Predict!

	Item 1	Item 2	...	Item m
User 1	1	5		4
User 2	5	1		?
...				
User n	2	5		?

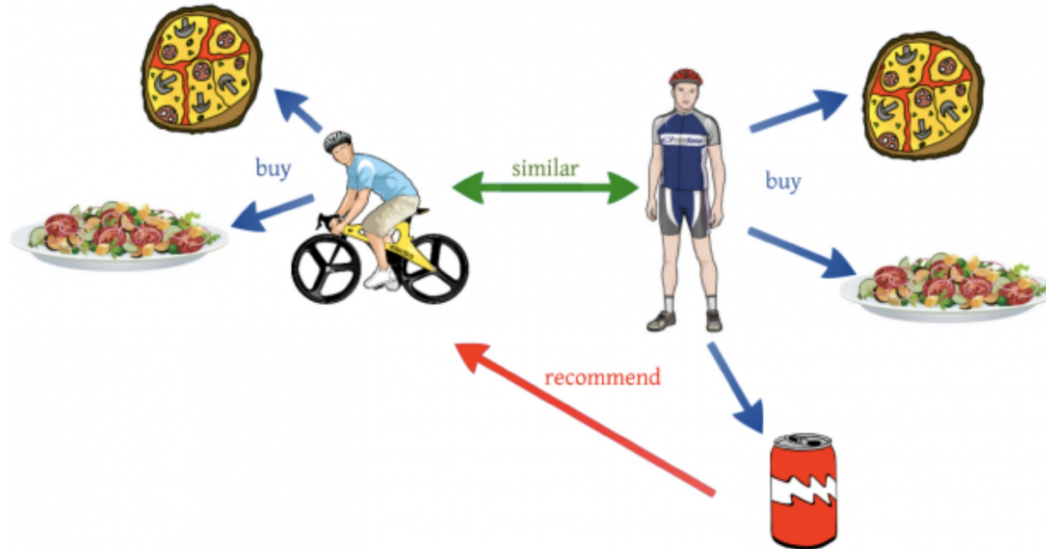


- ... y qué tan buena es mi predicción?

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{r}_{ui} - r_{ui})^2}{n}}$$

Filtrado Colaborativo basado en Usuarios

- **Objetivo:** buscar a usuarios similares y recomendar usando una suma ponderada con una métrica de similaridad



¿Cómo realiza las recomendaciones?

PSEUDO CÓDIGO.

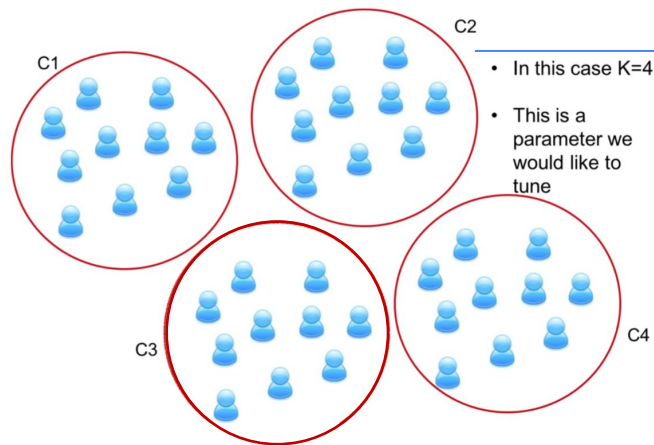
1. Busca N usuarios más similares y obtiene un score de similaridad.
2. Recorre ítems de estos usuarios que no hayan sido consumidos por el usuario objetivo.
3. Multiplica el rating de estos ítems por el score de similaridad con el usuario y guardarlos en una lista.
4. Ordena de mayor a menor puntaje ponderado.
5. Recomendando los top N.

Limitaciones del Filtrado Colaborativo

- **Exactitud:** mientras más **K** vecinos considere menor incertidumbre y mayor exactitud.
- **Escalabilidad:** a mayor usuarios, mayor es el costo computacional de encontrar los **K** vecinos más cercanos. Complejidad de K-NN es $O(dnk)$
- **Dispersión:** el método no se ajusta con usuarios o ítems con pocos ratings.
- **Cold start problem:** nuevos usuarios e ítems en el catalogo.

Alternativas al Filtrado Colaborativo

- Filtrado Colaborativo basado en Usuarios con **Clusters**



- Tomar vecinos dentro del *cluster* de usuarios representándolos con los ratings que han dado.
- Usar algoritmo k-means para calcular obtener *clusters de usuarios*.
- Generar la recomendación basada en items de usuarios cercanos

¿De qué otra forma podríamos generar recomendaciones utilizando Filtrado Colaborativo?

Hint: buscar otra entidad para calcular similitudes

Filtrado Colaborativo Basado en Ítems

El filtrado colaborativo basado en ítems se basa en la premisa de que si a una persona le gusta un ítem, es probable que le gusten los ítems similares.

USER-BASED

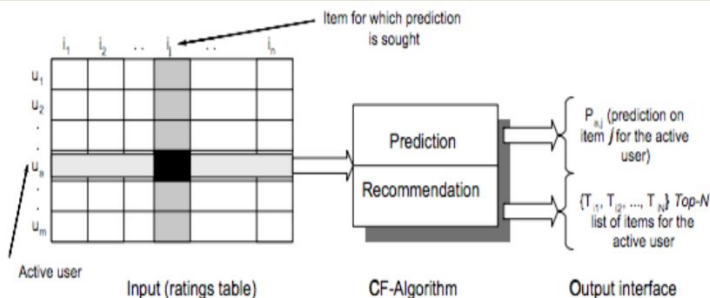


Figure 1: The Collaborative Filtering Process.

ITEM-BASED

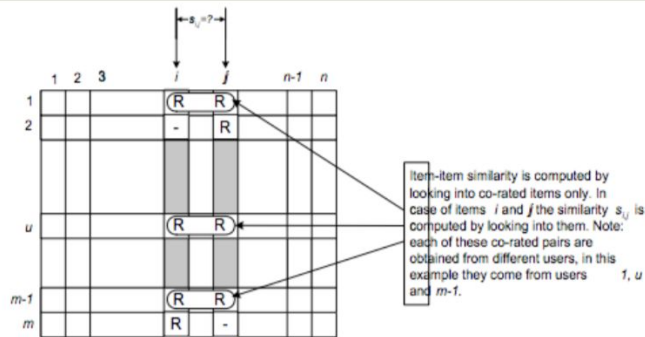


Figure 2: Isolation of the co-rated items and similarity computation

Cálculo de similaridad entre ítems y luego
recomendación ...

Fórmula de Similitud del Coseno

$$\text{sim}(i, j) = \frac{\sum_u r_{ui} \times r_{uj}}{\sqrt{\sum_u r_{ui}^2} \times \sqrt{\sum_u r_{uj}^2}}$$

- **sim(i, j)** es la similitud entre el ítem **i** y el ítem **j**.
- **r_{ui}, r_{uj}** calificaciones del usuario **u** para el ítem **i** e ítem **j**.
- **Sumatoria del denominador** multiplica ratings para usuarios que han calificado ítems **i** y **j** de manera conjunta.
- **Denominador** normaliza por la norma de las calificaciones de todos los usuarios para los ítems **i** y **j**.

Predicción de Calificación y Exclusión de Ítems Consumidos

$$\hat{r}_{ui} = \frac{\sum_{j \in N(i) \cap I'_u} \mathbf{sim}(i, j) \times r_{uj}}{\sum_{j \in N(i) \cap I'_u} |\mathbf{sim}(i, j)|}$$

$N(i) \cap I'_u$: ítems similares al ítem i que no han sido calificados por u .

$|\mathbf{sim}(i, j)|$: valor absoluto de similaridad entre ítem i y j

Acotando la Búsqueda de Ítems Similares

Para optimizar el proceso y acotar la búsqueda, puedes usar un umbral de similitud o limitar el número de ítems similares que consideras para las recomendaciones.

Ventajas de Filtrado Basado en Ítems

Escalabilidad: Funciona mejor para conjuntos de datos grandes ya que se precalcula la similitud entre ítems.

Estabilidad: Los ítems suelen ser más estables en sus características que los patrones de calificación de los usuarios.

Calidad de Recomendaciones: En muchos casos, se ha observado que las recomendaciones basadas en ítems son más precisas que las basadas en usuarios.

Análisis de Item Based CF

- [Sarwar et. al. \(2001\) Item-Based Collaborative Filtering Recommendation Algorithms](#)

- Métrica de evaluación

$$MAE = \frac{1}{N} \sum_{i=1}^N |r_{u,i} - \hat{r}_{u,i}|$$

- Resultados relevantes

- Efecto de la métrica de similaridad
- Sensitividad de la proporción *Training / Test*
- Tamaño de vecindario **K**
- Comparación con otros métodos

Análisis de Item Based CF: comparación de métricas de similitud.

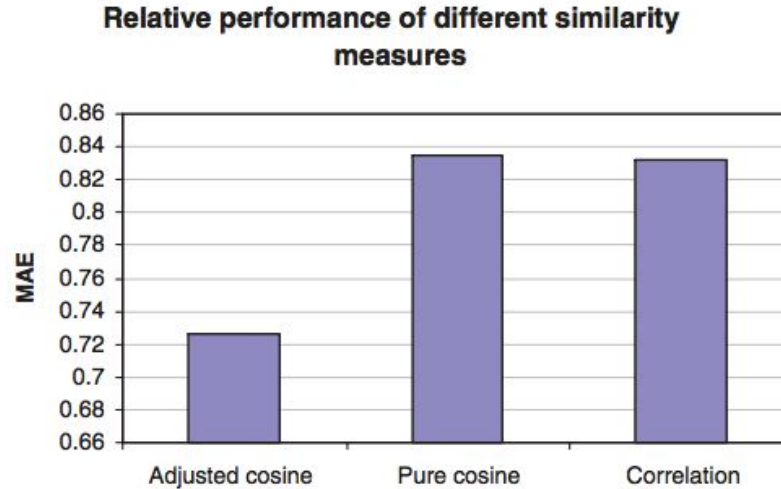


Figure 4: Impact of the similarity computation measure on item-based collaborative filtering algorithm.

Análisis de Item Based CF: comparación con otros métodos

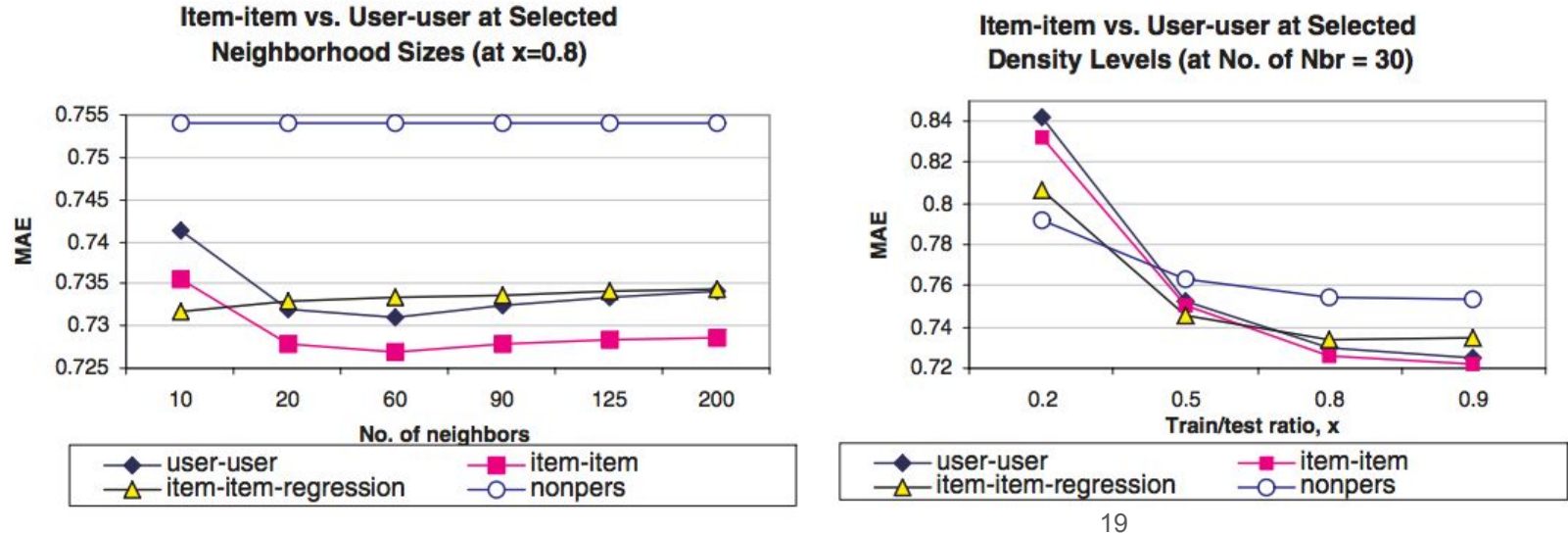


Figure 6: Comparison of prediction quality of *item-item* and *user-user* collaborative filtering algorithms. We compare prediction qualities at $x = 0.2, 0.5, 0.8$ and 0.9 .

Reflexión

¿Qué información captura el modelo de filtrado colaborativo basado en ítems?

¿Cómo realiza las recomendaciones?

¿Cómo realiza las recomendaciones?

PSEUDO CÓDIGO.

1. Busca items más similares (columnas) a los que consumió el usuario objetivo y obtiene un score de similaridad.
2. Recorre ítems que no hayan sido consumidos por el usuario objetivo.
3. Multiplica el rating de estos ítems por el score de similaridad con los items y guardarlos en una lista.
4. Ordena de mayor a menor puntaje ponderado.
5. Recomendando los top N.

Limitaciones de filtrado colaborativo basado en ítems.

- **Cold start:** Dificultad para recomendar a usuarios o ítems nuevos sin datos previos.
- **Tendencia a la popularidad:** Los ítems populares tienden a ser más recomendados, reduciendo la diversidad.
- **Esparsidad:** Con muchos ítems, solo unos pocos pueden tener suficientes calificaciones para determinar similitudes precisas.

Filtrado colaborativo (resumen)

Basada en usuario

