

Trabajo Sistema de Recomendación



Nuevas Tecnologías y Empresa

Nacho Santos jisantos @ ubu.es

Breve introducción a los Sistemas de recomendación

Problema

Amazon: ¿cómo aumentar las ventas?

Sistema de
recomendación

More Items to Consider



Puppeteer (PS3)
Sony
PlayStation 3
★★★★☆ (16)
£24.99



Kingdom Hearts 1.5: Standard Edition
Square Enix
PlayStation 3
★★★★☆ (44)
£19.99



Rayman Legends (PS3)
Ubisoft
PlayStation 3
★★★★☆ (23)
£19.99



Ratchet and Clank Nexus (PS3)
Sony
PlayStation 3
★★★★☆ (1)
£19.99



Ni No Kuni - Wrath of the White Witch
Namco Bandai
PlayStation 3
★★★★☆ (125)
£15.71



Tearaway (PlayStation Vita)
Sony
PlayStation Vita
£19.99

[View or edit your browsing history](#)

Related to Items You've Viewed



Disney Epic Mickey 2 - The Power of Two
PlayStation 3
★★★★☆ (21)
£15.20



F1 Race Stars (PS3)
Codemasters Limited
PlayStation 3
★★★★☆ (41)
£7.15



LEGO Batman: The Videogame (PS3)
Warner Bros. Interactive
PlayStation 3
★★★★☆ (58)
£15.48



Sonic Generations (PS3)
SEGA
PlayStation 3
★★★★☆ (56)
£12.07



LittleBigPlanet Karting (PS3)
Sony
PlayStation 3
★★★★☆ (54)
£11.99



Disney Universe (PS3)
PlayStation 3
★★★★☆ (15)
£14.99

Sistemas de recomendación

Sistemas de recomendación

Un **sistema de recomendación** filtra la información y muestra aquellos datos que potencialmente pueden interesar al usuario

Por ejemplo, muchas empresas e-commerce disponen de un **catálogo de productos** inmenso, y necesitan proponer aquellos productos que **a priori pueden interesar** al cliente

Cuanto más se acerquen a sus intereses mayores probabilidades de compra

Sistemas de recomendación

Sistemas de recomendación

- Filtros basados en contenidos
- Filtros colaborativos

Filtros basados en contenidos: analizan las características de los usuarios y el contenido de los productos para encontrar patrones que permitan hacer recomendaciones de nuevos productos

Filtros colaborativos: utilizan una matriz de utilidad, e.g. valoraciones (ratings), de los usuarios sobre un conjunto de productos para recomendar nuevos productos.



Sistemas de recomendación

Sistemas de recomendación

→ Filtros basados en contenidos

→ Filtros colaborativos

→ Basados en modelos

→ Basados en memoria

FC basados en modelos: utilizan los datos para ajustar modelos que después pueden ser utilizados para proponer recomendaciones (e.g. modelos de regresión como el visto en Gestión de la Información)

FC basados en memoria: utilizan los datos para definir similitudes entre usuarios y productos que serán utilizados para construir las recomendaciones (e.g. Amazon)

Sistemas de recomendación

Sistemas de recomendación

→ Filtros basados en contenidos

→ Filtros colaborativos

→ Basados en modelos

→ Basados en memoria

→ Basados en usuarios

→ Basados en productos

FC basados en usuarios:

La estimación del rating de un usuario u al producto i se basa en los ratings de otros **usuarios similares** a u al producto i

FC basados en productos: (e.g. Amazon)

La estimación del rating de un usuario u al producto i se basa en los ratings del usuario u a **productos similares** a i

Filtros colaborativos basados en memoria

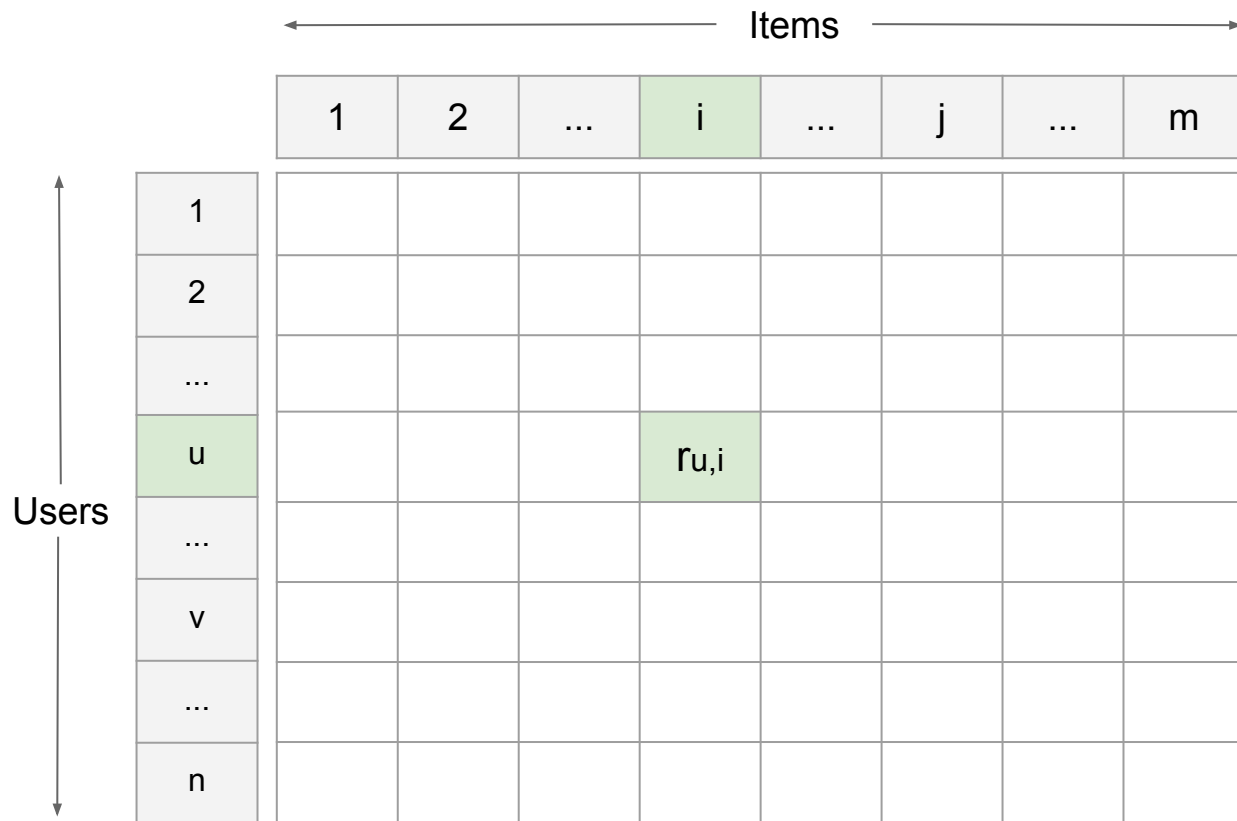
Filtros colaborativos basados en memoria

La **hipótesis** de un filtro colaborativo basado en memoria es que si dos usuarios han valorado de forma similar un conjunto de productos, lo harán también en otros productos similares (filtro basado en usuarios). De forma análoga podemos decir que si dos productos son similares, serán valorados de forma similar por un usuario (filtro basado en productos)

Los FC basados en memoria requieren:

1. Una **matriz de utilidad** (e.g. ratings)
2. Una definición de **similitud** (entre productos/usuarios)
3. Una definición de **vecindad** (subconjunto más próximo a un producto/usuario)
4. Una **regla para predecir** el rating en base al conjunto de productos/usuarios de la vecindad

1 Matriz de utilidad (e.g. ratings)



The diagram illustrates a utility matrix. A vertical double-headed arrow on the left is labeled "Users", and a horizontal double-headed arrow at the top is labeled "Items". The matrix is a grid with 8 rows and 8 columns. The first column contains user indices: 1, 2, ..., u, ..., v, ..., n. The first row contains item indices: 1, 2, ..., i, ..., j, ..., m. The cell at the intersection of row 'u' and column 'i' is highlighted in green and contains the rating $r_{u,i}$.

	1	2	...	i	...	j	...	m
1								
2								
...								
u				$r_{u,i}$				
...								
v								
...								
n								

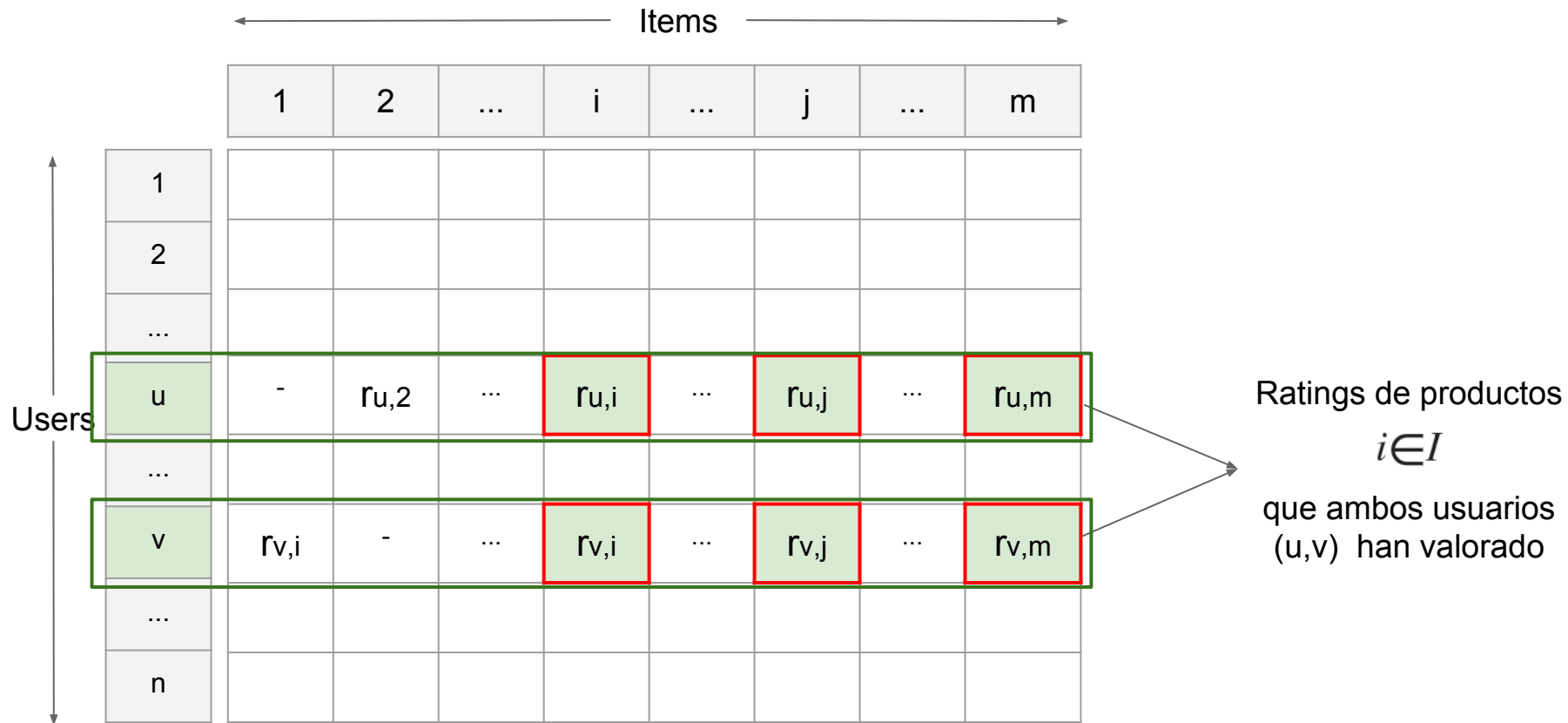
La matriz de utilidad representa una **función de utilidad**:

U: users x items \rightarrow rating

Ejemplos rating:

- Binario (1,0)
- Escala (1-5)

2 Definición de similitud: entre usuarios



2 Definición de similitud: entre usuarios

(1) Similitud del coseno (-1,1)

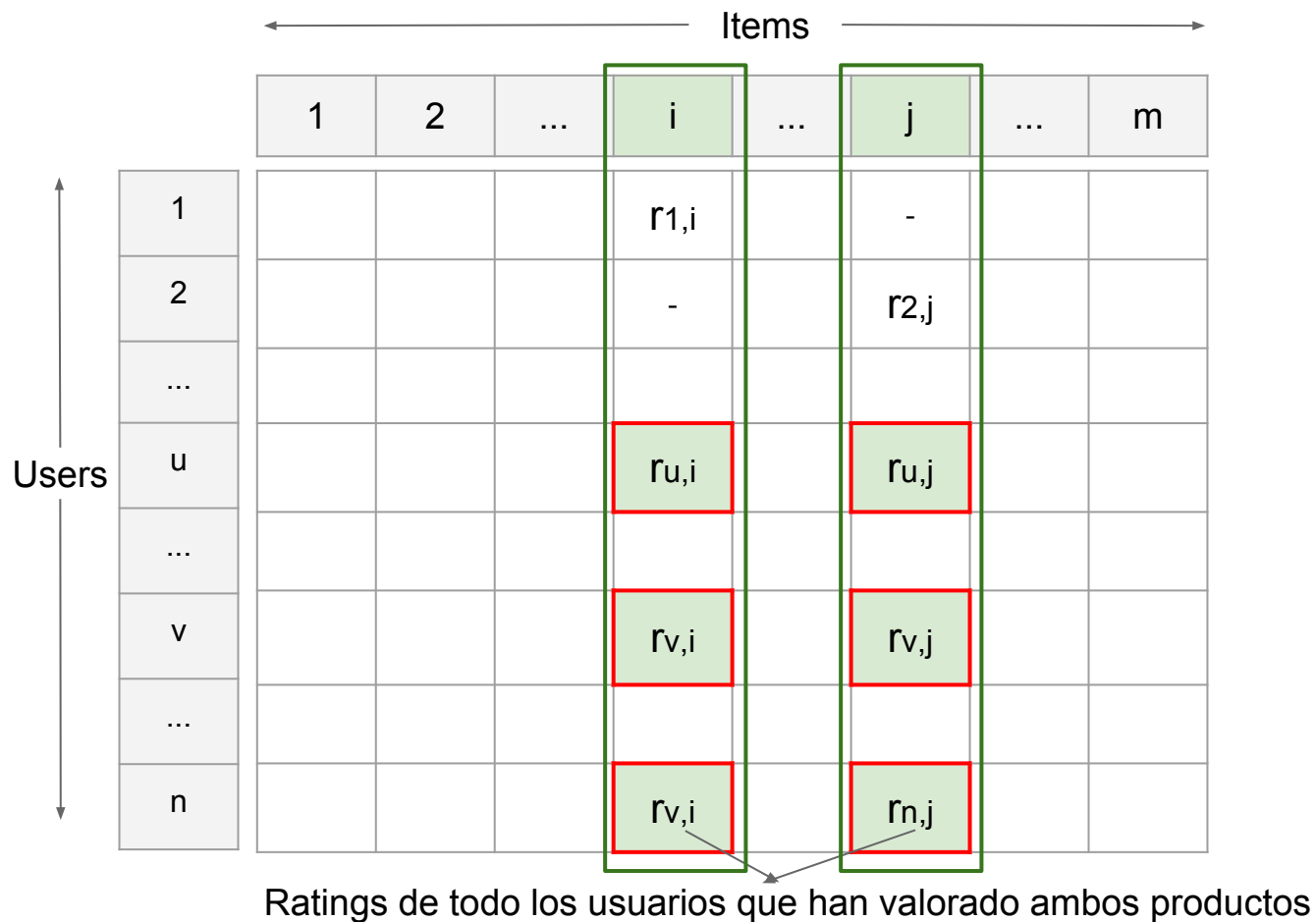
$$sim_{u,v} = \frac{\sum_{i \in I} r_{u,i} r_{v,i}}{\sqrt{\sum_{i \in I} r_{u,i}^2} \sqrt{\sum_{i \in I} r_{v,i}^2}}$$

(2) Coeficiente correlación de Pearson (-1,1)

$$sim_{u,v} = \frac{\sum_{i \in I} (r_{u,i} - \bar{r}_u) (r_{v,i} - \bar{r}_v)}{\sqrt{\sum_{i \in I} (r_{u,i} - \bar{r}_u)^2} \sqrt{\sum_{i \in I} (r_{v,i} - \bar{r}_v)^2}}$$

I productos
evaluados por **u,v**

2 Definición de similitud: entre productos



2 Definición de similitud: entre productos

(1) Similitud del coseno (-1,1)

$$sim_{i,j} = \frac{\sum_{u \in U} r_{u,i} r_{u,j}}{\sqrt{\sum_{u \in U} r_{u,i}^2} \sqrt{\sum_{u \in U} r_{u,j}^2}}$$

(2) Coeficiente correlación de Pearson (-1,1)

$$sim_{i,j} = \frac{\sum_{u \in U} (r_{u,i} - \bar{r}_i)(r_{u,j} - \bar{r}_j)}{\sqrt{\sum_{u \in U} (r_{u,i} - \bar{r}_i)^2} \sqrt{\sum_{u \in U} (r_{u,j} - \bar{r}_j)^2}}$$

U usuarios que
han evaluado **i, j**

3 Regla para estimar (predecir) un rating

Filtro colaborativo **basado en usuarios (user-based)**:

$$\hat{r}_{u,i} = \bar{r}_u + \frac{\sum_{v \in U} (r_{v,i} - \bar{r}_v) \text{sim}_{u,v}}{\sum_{v \in U} |\text{sim}_{u,v}|}$$

U usuarios* que han evaluado **i**

Filtro colaborativo **basado en productos (item-based)**:

$$\hat{r}_{u,i} = \frac{\sum_{j \in I} r_{u,j} \text{sim}_{i,j}}{\sum_{j \in I} |\text{sim}_{i,j}|}$$

I productos* evaluados por **u**

* Se suele hacer sobre un subconjunto de usuarios/productos: por ejemplo los **k vecinos más próximos**

Sistemas de recomendación

Hemos descrito en detalle los Filtros Colaborativos basados en memoria. Aunque son muy utilizados, presenta algunos inconvenientes (que también se dan en otros sistemas de recomendación), principalmente (Su & Khoshgoftaar, 2009):

- Muchos sistemas tienen **datos dispersos** lo que dificulta encontrar usuarios o productos similares (sobre los que basar una recomendación).
- Problema de “**cold start**”: dificultad de hacer recomendaciones a usuarios y/o productos nuevos (para los que no se dispone de información)
- Problemas de **escalabilidad** con el aumento del número de usuarios y/o productos
- Los **sinónimos** (dos o más referencias diferentes al mismo producto que sin embargo el filtro trata como productos distintos) introducen ruido en las recomendaciones
- “**Gray sheep**” y “**black sheep**”, cuando algunos usuarios tienen un patrón alejado del resto (o incluso único) el filtro colaborativo no hace buenas recomendaciones

Implementación en python

[Notebook Jupyter](#)

Propuesta de trabajo

Propuesta de trabajo

En grupos de 2-3 personas, se pide:

1 pto

1. Implementar un **filtro colaborativo “item-based”**
 - La descripción del filtro se encuentra en [Ekstrand \(2011\)](#) (Punto 2.3). La regla de predicción corresponde a la fórmula descrita anteriormente

1 pto

2. Implementar un **filtro colaborativo basado en la red bipartita** que forman los usuarios y los productos
 - La descripción del filtro se encuentra en [Sawant \(2013\)](#) (Apartado “Recommendation Power as Similarity”)
 - La regla de predicción es semejante a la del filtro colaborativo basado en usuarios:

$$\hat{r}_{u,i} = \overline{r}_u + \sum_{v \in U} (r_{v,i} - \overline{r}_v) sim_{u,v}$$

donde las medidas de similitud se denominan “**recommendation power**” y están basadas en la red bipartita de usuarios y productos

Propuesta de trabajo

- Vamos a utilizar los datos [Movielens](#) (100,000 ratings (1-5) de 943 usuarios sobre 1682 películas). Los datos se han organizado de la siguiente manera:
 - Un fichero csv {itemid,title} de películas [items.data](#)
 - Un fichero csv {userid,itemid,rating} con datos de ratings de **entrenamiento** [training.data](#)
 - Un fichero csv {userid,itemid,rating} con datos de ratings de **test** [test.data](#)
- Los datos de entrenamiento se utilizarán en el filtro colaborativo para estimar los ratings correspondientes a los datos de test (T), y se calculará una medida de error:
 - **Root Mean Squared Error (RMSE)**

$$RMSE = \sqrt{\frac{1}{|T|} \sum_{(u,b) \in T} (\hat{r}_{u,b} - r_{u,b})^2}$$

Propuesta de trabajo

- Se entregará un script de python/notebook de jupyter para cada filtro, con su implementación y el análisis del RMSE sobre los datos de test
- Cada filtro se evaluará con una nota máxima de 1 punto (sobre el total de 10 de la asignatura)
- Se dedicarán algunas de las horas de prácticas a trabajar en la implementación y resolver dudas.
- Existirá un foro en ubuvirtual para poder resolver las dudas de forma colaborativa
- La fecha de entrega será el día del examen **17 de junio**

Referencias

Ekstrand, M. D., Riedl, J. T., & Konstan, J. A. (2011). Collaborative filtering recommender systems. Foundations and Trends in Human-Computer Interaction, 4(2), 81-173. [[pdf](#)]

Sawant, S. (2013, December). Collaborative filtering using weighted bipartite graph projection: a recommendation system for yelp. In Proceedings of the CS224W: Social and Information Network Analysis Conference.[[pdf](#)]

Su, X., & Khoshgoftaar, T. M. (2009). A survey of collaborative filtering techniques. Advances in artificial intelligence, 2009, 4. [[pdf](#)]