

Práctica 2: Sistemas de recomendación

Enrique Mármol Campos

Universidad de Murcia
Máster en Big Data
ACUE

27 de febrero de 2024

Uno de los objetivos de empresas como Amazon o Netflix es dar recomendaciones al usuario de productos o contenidos que le puedan interesar. Para ello se basan en sugerencias personalizadas.

Uno de los objetivos de empresas como Amazon o Netflix es dar recomendaciones al usuario de productos o contenidos que le puedan interesar. Para ello se basan en sugerencias personalizadas.

Estas sugerencias son complicadas de conseguir, pues requieren de tener en cuenta multitud de factores para que la propuesta sea interesante al usuario. Este proceso hay que repetirlo varias veces ya que depende del usuario al que se quiera recomendar. Esto nos puede llevar a algunas estrategias de recomendación

Uno de los objetivos de empresas como Amazon o Netflix es dar recomendaciones al usuario de productos o contenidos que le puedan interesar. Para ello se basan en sugerencias personalizadas.

Estas sugerencias son complicadas de conseguir, pues requieren de tener en cuenta multitud de factores para que la propuesta sea interesante al usuario. Este proceso hay que repetirlo varias veces ya que depende del usuario al que se quiera recomendar. Esto nos puede llevar a algunas estrategias de recomendación

- Content-based.

Uno de los objetivos de empresas como Amazon o Netflix es dar recomendaciones al usuario de productos o contenidos que le puedan interesar. Para ello se basan en sugerencias personalizadas.

Estas sugerencias son complicadas de conseguir, pues requieren de tener en cuenta multitud de factores para que la propuesta sea interesante al usuario. Este proceso hay que repetirlo varias veces ya que depende del usuario al que se quiera recomendar. Esto nos puede llevar a algunas estrategias de recomendación

- Content-based.
- Collaborative

El principio general de los métodos basados en contenido (o cognitivos) es identificar las características comunes de los elementos que han recibido una calificación favorable de un usuario y luego recomendarle nuevos elementos que compartan esas características.

El principio general de los métodos basados en contenido (o cognitivos) es identificar las características comunes de los elementos que han recibido una calificación favorable de un usuario y luego recomendarle nuevos elementos que compartan esas características.

Nosotros partimos de un elemento $A = (a_i)_1^n$ con n atributos y de otro elemento $X = (x_i)_1^n$ del que queremos medir la similitud para poder recomendarlo.

- Cálculo de utilidad en base a la distancia:

$$\textit{Similitud}(A, X) = \frac{1}{1 + d(A, X)}$$

- Cálculo de utilidad en base a la distancia:

$$Similitud(A, X) = \frac{1}{1 + d(A, X)}$$

- Cálculo de utilidad en base a la similitud del coseno:

$$Similitud = \cos \theta = \frac{A \cdot X}{\|A\| \|X\|} = \frac{\sum_1^n a_i \cdot x_i}{\sqrt{\sum_1^n a_i^2} \sqrt{\sum_1^n x_i^2}}$$

- **Análisis de contenido limitado:**

- Información restringida sobre usuarios y elementos.
- Dificultad y coste en obtener detalles precisos (por ejemplo, en música o imágenes).
- El contenido disponible puede ser insuficiente para evaluar la calidad de los elementos.

- **Sobre-especialización:**

- Recomendación de elementos muy similares a los previamente gustados por el usuario.
- Limita la exploración de elementos novedosos pero interesantes.

Recomendación colaborativa.

La **recomendación colaborativa** es la más utilizada por empresas para sus recomendaciones, ya que se puede aplicar en multitud de productos y se entiende fácilmente.

Recomendación colaborativa.

La **recomendación colaborativa** es la más utilizada por empresas para sus recomendaciones, ya que se puede aplicar en multitud de productos y se entiende fácilmente.

En lugar de depender de la información del contenido, La recomendación colaborativa **utiliza la información de calificación de otros usuarios y elementos en el sistema.**

Recomendación colaborativa.

La **recomendación colaborativa** es la más utilizada por empresas para sus recomendaciones, ya que se puede aplicar en multitud de productos y se entiende fácilmente.

En lugar de depender de la información del contenido, La recomendación colaborativa **utiliza la información de calificación de otros usuarios y elementos en el sistema.**

La idea clave es que la calificación de un usuario objetivo para un nuevo elemento probablemente sea similar a la calificación de otro usuario si ambos han calificado otros elementos de manera similar.

Los enfoques colaborativos superan algunas de las limitaciones de los sistemas basados en contenido:

Recomendación colaborativa.

Los enfoques colaborativos superan algunas de las limitaciones de los sistemas basados en contenido:

- Las recomendaciones colaborativas se basan en la calidad de los elementos según la evaluación de los pares, en lugar de depender del contenido que puede ser un mal indicador de calidad.

Recomendación colaborativa.

Los enfoques colaborativos superan algunas de las limitaciones de los sistemas basados en contenido:

- Las recomendaciones colaborativas se basan en la calidad de los elementos según la evaluación de los pares, en lugar de depender del contenido que puede ser un mal indicador de calidad.
- A diferencia de los sistemas basados en contenido, los de filtrado colaborativo pueden recomendar elementos con contenido muy diferente

Recomendación colaborativa.

Los enfoques de filtrado colaborativo pueden agruparse en dos clases generales: métodos basado en memoria y métodos basados en modelos:

Los enfoques de filtrado colaborativo pueden agruparse en dos clases generales: métodos basado en memoria y métodos basados en modelos:

- **Basados en memoria:** las calificaciones de usuario-elemento almacenadas en el sistema se utilizan directamente para predecir calificaciones de nuevos elementos.

Los enfoques de filtrado colaborativo pueden agruparse en dos clases generales: métodos basado en memoria y métodos basados en modelos:

- **Basados en memoria:** las calificaciones de usuario-elemento almacenadas en el sistema se utilizan directamente para predecir calificaciones de nuevos elementos.
- **Basados en modelo:** utilizan las calificaciones almacenadas directamente para entrenar modelos que luego se emplean para predecir nuevas calificaciones.

Recomendación basada en memoria.

- **Simplicidad:** Métodos intuitivos y fáciles de implementar, con solo el parámetro del número de vecinos por ajustar.
- **Justificabilidad:** Permiten explicar las recomendaciones mostrando ejemplos claros (ítems y calificaciones), lo que ayuda a la comprensión y fomenta la interacción del usuario.
- **Eficiencia:** No requieren entrenamientos costosos; el pre-cálculo de vecinos es rápido y consume poca memoria, facilitando respuestas casi instantáneas en sistemas a gran escala.
- **Estabilidad:** Resisten la incorporación constante de nuevos usuarios, ítems y calificaciones, actualizando solo las similitudes necesarias sin reentrenar el sistema.

Recomendación colaborativa basada en usuario.

Los métodos de recomendación basados en usuario predicen el valor r_{ui} de un usuario u y un ítem i usando las valoraciones de i de los usuarios similares.

Recomendación colaborativa basada en usuario.

Los métodos de recomendación basados en usuario predicen la valoración r_{ui} de un usuario u y un ítem i usando las valoraciones de i de los usuarios similares.

	Item1	Item2	Item3	Item4	Item5
Alice	5	3	4	4	?
User1	3	1	2	3	3
User2	4	3	4	3	5
User3	3	3	1	5	4
User4	1	5	5	2	1

Recomendación colaborativa basada en usuario.

$$\hat{r}_{ui} = \frac{1}{N(u)} \sum_{v \in N(u)} r_{vi}$$

Recomendación colaborativa basada en usuario.

$$\hat{r}_{ui} = \frac{1}{N(u)} \sum_{v \in N(u)} r_{vi}$$

$$\hat{r}_{ui} = \frac{1}{\sum |w_{uv}|} \sum_{v \in N(u)} w_{uv} r_{vi}$$

Recomendación colaborativa basada en usuario.

$$\hat{r}_{ui} = \frac{1}{N(u)} \sum_{v \in N(u)} r_{vi}$$

$$\hat{r}_{ui} = \frac{1}{\sum |w_{uv}|} \sum_{v \in N(u)} w_{uv} r_{vi}$$

$$\hat{r}_{ui} = \bar{r}_u + \frac{1}{\sum |w_{uv}|} \sum_{v \in N(u)} w_{uv} (r_{vi} - \bar{r}_v)$$

Recomendación colaborativa basada en usuario.

$$\hat{r}_{ui} = \frac{1}{N(u)} \sum_{v \in N(u)} r_{vi}$$

$$\hat{r}_{ui} = \frac{1}{\sum |w_{uv}|} \sum_{v \in N(u)} w_{uv} r_{vi}$$

$$\hat{r}_{ui} = \bar{r}_u + \frac{1}{\sum |w_{uv}|} \sum_{v \in N(u)} w_{uv} (r_{vi} - \bar{r}_v)$$

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

Recomendación colaborativa basada en usuario.

Dada la matriz de calificaciones:

Usuario	The Matrix	Titanic	Die Hard	Forrest Gump	Wall-E
John	5	1	—	2	2
Lucy	1	5	2	5	5
Eric	2	?	3	5	4
Diane	4	3	5	3	—

Recomendación colaborativa basada en usuario.

Se calculan los promedios:

$$\bar{r}_{\text{John}} = \frac{5 + 1 + 2 + 2}{4} = 2.5,$$

$$\bar{r}_{\text{Lucy}} = \frac{1 + 5 + 2 + 5 + 5}{5} = 3.6,$$

$$\bar{r}_{\text{Eric}} = \frac{2 + 3 + 5 + 4}{4} = 3.5,$$

$$\bar{r}_{\text{Diane}} = \frac{4 + 3 + 5 + 3}{4} = 3.75.$$

Recomendación colaborativa basada en usuario.

Se calculan los promedios:

$$\bar{r}_{\text{John}} = \frac{5 + 1 + 2 + 2}{4} = 2.5,$$

$$\bar{r}_{\text{Lucy}} = \frac{1 + 5 + 2 + 5 + 5}{5} = 3.6,$$

$$\bar{r}_{\text{Eric}} = \frac{2 + 3 + 5 + 4}{4} = 3.5,$$

$$\bar{r}_{\text{Diane}} = \frac{4 + 3 + 5 + 3}{4} = 3.75.$$

Las similitudes de Pearson entre Eric y los otros usuarios son:

$$\text{sim}(\text{Eric}, \text{John}) \approx -0.837,$$

$$\text{sim}(\text{Eric}, \text{Lucy}) \approx 0.921,$$

$$\text{sim}(\text{Eric}, \text{Diane}) \approx -0.66.$$

Recomendación colaborativa basada en usuario.

Las contribuciones son:

$$\text{John: } -0.837 \times (1 - 2.5) \approx 1.2555,$$

$$\text{Lucy: } 0.921 \times (5 - 3.6) \approx 1.2894,$$

$$\text{Diane: } -0.66 \times (3 - 3.75) \approx 0.495.$$

Recomendación colaborativa basada en usuario.

Las contribuciones son:

$$\text{John: } -0.837 \times (1 - 2.5) \approx 1.2555,$$

$$\text{Lucy: } 0.921 \times (5 - 3.6) \approx 1.2894,$$

$$\text{Diane: } -0.66 \times (3 - 3.75) \approx 0.495.$$

Con ello, se tiene:

$$\text{Numerador} = 1.2555 + 1.2894 + 0.495 \approx 3.04,$$

$$\text{Denominador} = |-0.837| + |0.921| + |-0.66| \approx 0.837 + 0.921 + 0.66 \approx 2.418$$

Recomendación colaborativa basada en usuario.

Las contribuciones son:

$$\text{John: } -0.837 \times (1 - 2.5) \approx 1.2555,$$

$$\text{Lucy: } 0.921 \times (5 - 3.6) \approx 1.2894,$$

$$\text{Diane: } -0.66 \times (3 - 3.75) \approx 0.495.$$

Con ello, se tiene:

$$\text{Numerador} = 1.2555 + 1.2894 + 0.495 \approx 3.04,$$

$$\text{Denominador} = |-0.837| + |0.921| + |-0.66| \approx 0.837 + 0.921 + 0.66 \approx 2.418$$

Finalmente:

$$\hat{r}_{\text{Eric, Titanic}} \approx 3.5 + \frac{3.04}{2.418} \approx 3.5 + 1.257 \approx 4.76.$$

Recomendación colaborativa basada en usuario.

Las contribuciones son:

$$\text{John: } -0.837 \times (1 - 2.5) \approx 1.2555,$$

$$\text{Lucy: } 0.921 \times (5 - 3.6) \approx 1.2894,$$

$$\text{Diane: } -0.66 \times (3 - 3.75) \approx 0.495.$$

Con ello, se tiene:

$$\text{Numerador} = 1.2555 + 1.2894 + 0.495 \approx 3.04,$$

$$\text{Denominador} = |-0.837| + |0.921| + |-0.66| \approx 0.837 + 0.921 + 0.66 \approx 2.418$$

Finalmente:

$$\hat{r}_{\text{Eric, Titanic}} \approx 3.5 + \frac{3.04}{2.418} \approx 3.5 + 1.257 \approx 4.76.$$

$r_{\text{Eric, Titanic}} \approx 4.76.$

Recomendación colaborativa basada en items

Mientras que los métodos basados recaen mas en la opinión de usuarios parecidos para predecir las valoraciones, los basados en items miran las valoraciones de items similares

Recomendación colaborativa basada en items

Mientras que los métodos basados recaen mas en la opinión de usuarios parecidos para predecir las valoraciones, los basados en items miran las valoraciones de items similares

	Item1	Item2	Item3	Item4	Item5
Alice	5	3	4	4	?
User1	3	1	2	3	3
User2	4	3	4	3	5
User3	3	3	1	5	4
User4	1	5	5	2	1

Recomendación colaborativa basada en items

Mientras que los métodos basados recaen mas en la opinión de usuarios parecidos para predecir las valoraciones, los basados en items miran las valoraciones de items similares

	Item1	Item2	Item3	Item4	Item5
Alice	5	3	4	4	?
User1	3	1	2	3	3
User2	4	3	4	3	5
User3	3	3	1	5	4
User4	1	5	5	2	1

$$pred(a, p) = \frac{\sum_{i \in I(a)} sim(i, p) \cdot r_{a,i}}{\sum_{i \in I(a)} sim(i, p)}$$

Recomendación colaborativa basada en items

Usuario	The Matrix	Titanic	Die Hard	Forrest Gump	Wall-E
John	5	1	—	2	2
Lucy	1	5	2	5	5
Eric	2	?	3	5	4
Diane	4	3	5	3	—

Recomendación colaborativa basada en items

$$\text{sim}(\text{Matrix}, \text{Titanic}) = \frac{5 \cdot 1 + 1 \cdot 5 + 4 \cdot 3}{\sqrt{5^2 + 1^2 + 4^2} \sqrt{1^2 + 5^2 + 3^2}} \approx 0.573,$$

$$\text{sim}(\text{Die Hard}, \text{Titanic}) = \frac{2 \cdot 5 + 5 \cdot 3}{\sqrt{2^2 + 5^2} \sqrt{5^2 + 3^2}} \approx 0.795,$$

$$\text{sim}(\text{Forrest Gump}, \text{Titanic}) = \frac{2 \cdot 1 + 5 \cdot 5 + 3 \cdot 3}{\sqrt{2^2 + 5^2 + 3^2} \sqrt{1^2 + 5^2 + 3^2}} \approx 0.988,$$

$$\text{sim}(\text{Wall-E}, \text{Titanic}) = \frac{2 \cdot 1 + 5 \cdot 5}{\sqrt{2^2 + 5^2} \sqrt{1^2 + 5^2}} \approx 0.983.$$

$$\text{pred}(\text{Eric}, \text{Titanic}) = \frac{12.403}{3.339} \approx 3.71.$$

$r_{\text{Eric}, \text{Titanic}} \approx 3.71$

La elección entre modelos de recomendación basados en usuario e ítems depende de las prioridades del sistema:

- Si se busca adaptabilidad rápida y un análisis directo de la comunidad, los enfoques basados en usuario pueden ser más adecuados, aunque pueden presentar problemas de escalabilidad y estabilidad.
- Por otro lado, los métodos basados en ítems ofrecen mayor escalabilidad y estabilidad, con una capacidad de justificabilidad relativamente clara, pero podrían limitar la serendipia si no se integran mecanismos que fomenten la diversidad.

Recomendacion basada en modelo

En el enfoque basado en modelos, se utilizan modelos de aprendizaje automático para predecir y clasificar las interacciones entre los usuarios y los elementos con los que aún no han interactuado.

En el enfoque basado en modelos, se utilizan modelos de aprendizaje automático para predecir y clasificar las interacciones entre los usuarios y los elementos con los que aún no han interactuado.

- Factorización matricial

En el enfoque basado en modelos, se utilizan modelos de aprendizaje automático para predecir y clasificar las interacciones entre los usuarios y los elementos con los que aún no han interactuado.

- Factorización matricial
- Deep learning

En el enfoque basado en modelos, se utilizan modelos de aprendizaje automático para predecir y clasificar las interacciones entre los usuarios y los elementos con los que aún no han interactuado.

- Factorización matricial
- Deep learning
- Clustering

Factorización matricial

La factorización matricial es una forma de generar características latentes al multiplicar dos tipos diferentes de entidades. Los sistemas de recomendación es la aplicación de la factorización matricial para identificar la relación entre las valoraciones de los items y las de los usuarios.

	Item			
	W	X	Y	Z
A		4.5	2.0	
B	4.0		3.5	
C		5.0		2.0
D		3.5	4.0	1.0

Rating Matrix

=

A	1.2	0.8
B	1.4	0.9
C	1.5	1.0
D	1.2	0.8

User Matrix

X

	W	X	Y	Z
	1.5	1.2	1.0	0.8
	1.7	0.6	1.1	0.4

Item Matrix

Tutorial de “recommenderlab” con el dataset Jester5k.

- Descarga el tutorial sobre la librería “recommenderlab”. Enlace: <https://cran.rproject.org/web/packages/recommenderlab/vignettes/recommenderlab.pdf>
- Sección 5.4. Inspecciona las propiedades del dataset Jester5k.
- Crea un imagen de las 100 primeras filas y columnas de la matriz con el comando: `image(Jester5k[1:100,1:100])`. ¿Qué usuarios son los que aportan más ratings? ¿Qué chistes han sido más valorados?
- Sección 5.5. Crea un recomendador siguiendo las indicaciones del tutorial. Crea otra versión del recomendador sólo con los 100 primeros usuarios y vuelve a predecir para los usuarios 1001 y 1002. ¿Detectas diferencias en las recomendaciones?

- Secciones 5.6 y 5.7. Evalúa las recomendaciones siguiendo el tutorial. ¿Qué diferencia hay entre la evaluación de la sección 5.6 y la 5.7? ¿Cuál te parece más apropiada para un sistema de recomendación?
- Sección 5.8. Compara los diferentes métodos de recomendación en esta sección. ¿Cómo quedaría el algoritmo de recomendación ítem-based en comparación con los utilizados en esta sección? ¿Cambian los resultados si modificamos el particionamiento utilizado en el procedimiento de evaluación?