



Universidad
de La Laguna

Sistemas de recomendación. Métodos de filtrado colaborativo

Gestión del conocimiento de las
organizaciones

Mariajose Zuloeta Brito
Samuel Lorenzo Sánchez
Adrián Lima García
Andrés Hernández Ortega



Introducción

En esta práctica, implementaremos un sistema de recomendación utilizando el método de filtrado colaborativo. Para comprender su funcionamiento, nos apoyaremos en las diapositivas del curso y desarrollaremos un software con tecnologías web como HTML, CSS3 y Vue. El sistema incluirá los siguientes elementos:

- Fichero de entrada que contendrá la matriz de utilidad.
- Métricas de similitud. Se utilizarán tres opciones:
 - Correlación de Pearson.
 - Distancia coseno.
 - Distancia euclídea.
- Predicción: se calculará la predicción de valores faltantes en la matriz, basada en un número determinado de vecinos, aplicando tanto la predicción simple como la basada en la diferencia con la media.

Como resultado, el sistema deberá generar:

1. Una matriz con las predicciones de los elementos faltantes.
2. Una matriz de similitudes entre todos los pares de usuarios.
3. La selección de los vecinos y el cálculo detallado de cada predicción en la matriz de utilidad.



Análisis

Los ficheros con la información obtenida en las pruebas se encuentran en el siguiente directorio compartido: [Pruebas](#)

Ejemplo tipo para una matriz de 5 usuarios

```
0.00
5.00
5.00 3.00 4.00 4.00 -
3.00 1.00 2.00 3.00 3.00
4.00 3.00 4.00 3.00 5.00
3.00 3.00 1.00 5.00 4.00
1.00 5.00 5.00 2.00 1.00
```

Resultado1_1.txt

Número de vecinos: 2

Métrica: Correlación de Pearson

Tipo de predicción: Simple

Matriz de Similitud:

```
Usuario 1: 1.0000 | 0.8528 | 0.7071 | 0.0000 | -0.7921
Usuario 2: 0.8528 | 1.0000 | 0.4677 | 0.4900 | -0.9001
Usuario 3: 0.7071 | 0.4677 | 1.0000 | -0.1612 | -0.4666
Usuario 4: 0.0000 | 0.4900 | -0.1612 | 1.0000 | -0.6415
Usuario 5: -0.7921 | -0.9001 | -0.4666 | -0.6415 | 1.0000
```

Matriz de Utilidad con Predicciones:

```
Usuario 1: 5.0000 | 3.0000 | 4.0000 | 4.0000 | 4.8720
Usuario 2: 3.0000 | 1.0000 | 2.0000 | 3.0000 | 3.0000
Usuario 3: 4.0000 | 3.0000 | 4.0000 | 3.0000 | 5.0000
Usuario 4: 3.0000 | 3.0000 | 1.0000 | 5.0000 | 4.0000
Usuario 5: 1.0000 | 5.0000 | 5.0000 | 2.0000 | 1.0000
```

Conjunto de pares de usuarios:

Usuario 1 con el ítem 5

Vecino 2 (Similitud: 0.8528)

Vecino 3 (Similitud: 0.7071)

Media del usuario 1 es 4.0000

$$\text{SUMATORIO}(\text{similitud}(a,b) * \text{puntuación}(b) - \text{mediaPuntuaciones}(b)) = 1.3602$$

$$\text{SUMATORIO}(\text{similitud}(a,b)) = 1.5599$$



Siendo la predicción $1.3602 / 1.5599 = 4.8720$ para el ítem: 5

Conclusiones

Los usuarios 1 y 2 tienen la mayor similitud (0.8528), mientras que el usuario 1 y el 5 tienen la menor (-0.7921), lo cual implica que el usuario 5 tiene un perfil de preferencias casi opuesto al del usuario 1.

Para predecir la puntuación del usuario 1 en el ítem 5 (con predicción simple) se emplean los dos vecinos más similares, que en este caso son los usuarios 2 y 3.

Con una media de puntuación del usuario 1 de 4.0000 y una combinación ponderada de desviaciones en relación a las medias de los vecinos (con similitudes 0.8528 y 0.7071), se obtiene un valor de predicción de 4.8720 para el ítem 5. Esto sugiere que es probable que al usuario 1 le guste el ítem 5, con una predicción cercana al límite superior de puntuación posible.

La predicción basada en los vecinos posibilita estimar con cierto nivel de precisión la calificación de ítems no puntuados, asumiendo que existe una coherencia suficiente en los patrones de preferencia de los usuarios cercanos.

El empleo de la Correlación de Pearson, aunque puede resultar efectivo para capturar similitudes en patrones, no aborda posibles sesgos derivados de variaciones en los niveles de puntuación absolutos entre usuarios.

El filtrado colaborativo basado en similitudes permite predecir con éxito calificaciones de ítems no puntuados por un usuario a partir de sus vecinos más similares.

Resultado1_2.txt

Número de vecinos: 2

Métrica: Distancia de Coseno

Tipo de predicción: Diferencia con la Media

Matriz de Similitud:

Usuario 1:	1.0000		0.9753		0.9922		0.8907		0.7967
Usuario 2:	0.9753		1.0000		0.9594		0.9357		0.6378
Usuario 3:	0.9922		0.9594		1.0000		0.8944		0.7715
Usuario 4:	0.8907		0.9357		0.8944		1.0000		0.6383
Usuario 5:	0.7967		0.6378		0.7715		0.6383		1.0000

Matriz de Utilidad con Predicciones:

Usuario 1:	5.0000		3.0000		4.0000		4.0000		4.8720
------------	--------	--	--------	--	--------	--	--------	--	--------



```
Usuario 2: 3.0000 | 1.0000 | 2.0000 | 3.0000 | 3.0000
Usuario 3: 4.0000 | 3.0000 | 4.0000 | 3.0000 | 5.0000
Usuario 4: 3.0000 | 3.0000 | 1.0000 | 5.0000 | 4.0000
Usuario 5: 1.0000 | 5.0000 | 5.0000 | 2.0000 | 1.0000
```

Conjunto de pares de usuarios:

Usuario 1 con el ítem 5

Vecino 2 (Similitud: 0.8528)

Vecino 3 (Similitud: 0.7071)

Media del usuario 1 es 4.0000

$$\text{SUMATORIO(similitud(a,b) * puntuación(b) - mediaPuntuaciones(b))} = 1.3602$$

$$\text{SUMATORIO(similitud(a,b))} = 1.5599$$

Siendo la predicción $1.3602 / 1.5599 = 4.8720$ para el ítem: 5

Conclusiones

La matriz de similitud con la Distancia de Coseno muestra una mayor similitud promedio entre los usuarios en comparación con la métrica de Correlación de Pearson.

Los coeficientes de similitud entre los usuarios son generalmente altos (cerca de 1.0000), lo que sugiere una gran proximidad en los vectores de puntuación de los usuarios.

En el caso de los usuarios 1 y 3 se presenta una similitud de 0.9922, lo que implica que sus preferencias son casi idénticas en cuanto a dirección, aunque no necesariamente en cuanto a magnitud.

La predicción se realiza empleando la media del usuario 1 y las desviaciones de las puntuaciones de los vecinos en relación a sus propias medias.

La fórmula toma la similitud ponderada de las desviaciones (en relación a las medias de los vecinos) para ajustar la predicción. En este caso, la media del usuario 1 es 4.0000, y el cálculo arroja una predicción de 4.8720 para el ítem 5.

El resultado es el mismo valor de predicción para el ítem 5 que se obtuvo con la Correlación de Pearson, sugiriendo una consistencia en la predicción debido a los altos niveles de similitud entre usuarios con la métrica de Coseno.

La Distancia de Coseno permite medir la similitud de dirección en las preferencias, sin depender de los valores absolutos, lo cual es útil para detectar afinidades en patrones incluso cuando los usuarios puntúan en diferentes escalas.



Resultado1_3.txt

Número de vecinos: 4

Métrica: Distancia Euclídea

Tipo de predicción: Diferencia con la Media

Matriz de Similitud:

Usuario 1:	0.0000		3.6056		1.4142		3.7417		5.0000
Usuario 2:	3.6056		0.0000		3.6056		3.1623		5.8310
Usuario 3:	1.4142		3.6056		0.0000		3.8730		5.5678
Usuario 4:	3.7417		3.1623		3.8730		0.0000		6.4807
Usuario 5:	5.0000		5.8310		5.5678		6.4807		0.0000

Matriz de Utilidad con Predicciones:

Usuario 1:	5.0000		3.0000		4.0000		4.0000		3.8440
Usuario 2:	3.0000		1.0000		2.0000		3.0000		3.0000
Usuario 3:	4.0000		3.0000		4.0000		3.0000		5.0000
Usuario 4:	3.0000		3.0000		1.0000		5.0000		4.0000
Usuario 5:	1.0000		5.0000		5.0000		2.0000		1.0000

Conjunto de pares de usuarios:

Usuario 1 con el ítem 5

Vecino 5 (Similitud: 5.0000)

Vecino 4 (Similitud: 3.7417)

Vecino 2 (Similitud: 3.6056)

Vecino 3 (Similitud: 1.4142)

Media del usuario 1 es 4.0000

$$\text{SUMATORIO}(\text{similitud}(a,b) * \text{puntuación}(b) - \text{mediaPuntuaciones}(b)) = -2.1463$$

$$\text{SUMATORIO}(\text{similitud}(a,b)) = 13.7614$$

Siendo la predicción $-2.1463 / 13.7614 = 3.8440$ para el ítem: 5

Conclusiones

La matriz de similitud muestra la Distancia Euclídea entre cada par de usuarios. A diferencia de la Correlación de Pearson o la Distancia Coseno, donde la similitud aumenta al acercarse a 1, en este caso, la similitud es inversamente proporcional: una distancia menor indica una mayor similitud entre usuarios, siendo cero el valor que indica una similitud exacta, dándose en la matriz de similitud con los usuarios consigo mismos.



Los usuarios más similares al usuario 1 son los usuarios 3 (1.4142) y 2 (3.6056), mientras que el usuario 5 tiene la mayor distancia con el usuario 1 (5.0000). La selección de los 4 vecinos incluye tanto usuarios con distancias relativamente cortas (más similares) como usuarios más distantes, lo que podría introducir variabilidad en la predicción.

La predicción para el ítem 5 del usuario 1 emplea la media de las puntuaciones del propio usuario (4.0000) y ajusta el valor con base en la desviación de los vecinos seleccionados.

En este caso, se calcula una predicción de 3.8440 para el ítem 5, lo cual indica una expectativa de menor afinidad del usuario 1 hacia el ítem 5 en comparación con la predicción obtenida con la métrica de Coseno o la Correlación de Pearson.

La influencia de vecinos distantes (usuarios 4 y 5) con patrones de puntuación menos similares puede estar generando un sesgo, que reduce el valor de la predicción para el ítem 5, comparado con la puntuación media del usuario.

En este caso, los usuarios más cercanos (como el usuario 3) proporcionan una similitud elevada, mientras que los usuarios con distancias mayores (usuarios 4 y 5) tienden a reducir la precisión de la predicción en este tipo de recomendación, ya que sus puntuaciones son menos correlativas con las del usuario 1.

La distancia euclídea puede ser útil para identificar los vecinos más próximos en las puntuaciones absolutas, no obstante puede provocar cierto sesgo a la hora de no diferenciar una distancia que tiende hacia una dirección (puntuaciones más altas o bien puntuaciones más bajas). Sin embargo, es una métrica que posee operaciones de cálculo más simples, lo que puede ser atractivo para ciertas situaciones.



Ejemplo tipo para una matriz de 5 usuarios

```
0.00
5.00
- - - - -
3.00 1.00 - 3.00 3.00
4.00 3.00 - 3.00 5.00
3.00 3.00 - 5.00 4.00
1.00 5.00 - 2.00 -
```

Resultado2_1.txt

Número de vecinos: 3

Métrica: Correlación de Pearson

Tipo de predicción: Simple

Número de vecinos: 3

Métrica: Correlación de Pearson

Tipo de predicción: Simple

Matriz de Similitud:

```
Usuario 1: 1.0000 | N/A | 0.8783 | -0.9707
Usuario 2: N/A | 1.0000 | N/A | N/A
Usuario 3: 0.8783 | N/A | 1.0000 | -0.7206
Usuario 4: -0.9707 | N/A | -0.7206 | 1.0000
```

Matriz de Utilidad con Predicciones:

```
Usuario 1: 3.0000 | 2.0000 | 3.0000 | 3.0000
Usuario 2: 3.0000 | 3.0000 | 3.0000 | 3.0000
Usuario 3: 3.0000 | 1.0000 | 5.0000 | 4.0000
Usuario 4: 1.0000 | 5.0000 | 2.0000 | -3.4260
```

Conjunto de pares de usuarios:

Usuario 4 con el ítem 4

Vecino 3 (Similitud: -0.7206)

Vecino 1 (Similitud: -0.9707)

$SUMATORIO(similitud(a,b) * puntuacion(b,i)) = -5.7945$

$abs(SUMATORIO(similitud(a,b))) = 1.6913$

Siendo la predicción $-5.7945 / 1.6913 = -3.4260$ para el ítem: 4



Conclusiones

Como se puede observar en el ejemplo, no se incluyen valoraciones en el usuario 1 y ni en el ítem 3. Para lo cual se obvian.

Ejemplo tipo para una matriz de 10 usuarios

`Matriz3.txt`

`Resultado3_1.txt`

Número de vecinos: 2

Métrica: Correlación de Pearson

Tipo de predicción: Predicción Simple

`Resultado3_2.txt`

Número de vecinos: 4

Métrica: Distancia coseno

Tipo de predicción: Simple

`Resultado3_3.txt`

Número de vecinos: 2

Métrica: Distancia Euclídiana

Tipo de predicción: Diferencia con la media



Ejemplo tipo para una matriz de 25 usuarios

[Matriz5.txt](#)

[Resultado5_1.txt](#)

Número de vecinos: 3

Métrica: Correlación de Pearson

Tipo de predicción: Simple

[Resultado5_2.txt](#)

Número de vecinos: 5

Métrica: Distancia de Coseno

Tipo de predicción: Diferencia con la Media

[Resultado5_3.txt](#)

Número de vecinos: 6

Métrica: Distancia Euclidiana

Tipo de predicción: Diferencia con la Media



Ejemplo tipo para una matriz de 50 usuarios

[Matriz 7.txt](#)

[Resultado7_1.txt](#)

Número de vecinos: 6

Métrica: Correlación de Pearson

Tipo de predicción: Predicción Simple

[Resultado7_2.txt](#)

Número de vecinos: 10

Métrica: Distancia coseno

Tipo de predicción: Simple

[Resultado7_3.txt](#)

Número de vecinos: 14

Métrica: Distancia Euclidiana

Tipo de predicción: Diferencia con la media

Ejemplo tipo para una matriz de 100 usuarios

[Matriz8.txt](#)

[Resultado8_1.txt](#)

Número de vecinos: 10

Métrica: Correlación de Pearson

Tipo de predicción: Predicción Simple



Resultado8_2.txt

Número de vecinos: 2

Métrica: Distancia coseno

Tipo de predicción: Diferencia con la Media

Resultado8_3.txt

Número de vecinos: 20

Métrica: Distancia Euclidiana

Tipo de predicción: Diferencia con la media

Conclusiones

En esta práctica, implementamos un sistema de recomendación utilizando el método de filtrado colaborativo. Para entender su funcionamiento, nos apoyamos en las diapositivas del curso y desarrollamos un software con tecnologías web como HTML, CSS3 y Vue. El sistema incluye un fichero de entrada que contenía la matriz de utilidad, así como métricas de similitud basadas en tres opciones: la correlación de Pearson, la distancia coseno y la distancia euclídea. Además, calculamos la predicción de valores faltantes en la matriz, basada en un número determinado de vecinos, aplicando tanto la predicción simple como la basada en la diferencia con la media. Como resultado, el sistema genera una matriz con las predicciones de los elementos faltantes, una matriz de similitudes entre todos los pares de usuarios y la selección de los vecinos, junto con el cálculo detallado de cada predicción en la matriz de utilidad.

Durante el desarrollo, enfrentamos varias dificultades. En particular, el uso de componentes en Vue nos costó emplearlos correctamente, especialmente en la actualización de datos entre componentes. Además, al leer los datos desde archivos, algunos espacios en blanco generaron problemas en la interpretación de las matrices, lo cual afectó al cálculo de las similitudes devolviendo N/A en algunas filas y predicciones. Por último, trabajar en equipo en diferentes ramas en GitHub exigió una coordinación constante y una adecuada organización, ya que en ocasiones surgieron conflictos en el código que requirieron revisiones adicionales.