



**Escuela Superior
de Ingeniería y Tecnología**
Universidad de La Laguna

Sistemas de Recomendación

Método de filtrado colaborativo

Gestión del conocimiento

Miguel Alejandro Martín Reyes
alu0101209777@ull.edu.es



Índice:

Introducción	2
Formato de la matriz	2
Matriz de similitud	4
Selección de vecinos	5
Predicciones	7
Descargar informe	8
Ejemplos	9
Matriz 5 x 10	9
Matriz 10 x 25	10



Introducción

A continuación se describe la implementación web de un sistema de recomendación siguiendo el método de filtrado colaborativo.

Dicho sistema permite aplicar el método sobre matrices en las que las filas representan los usuarios y las columnas las valoraciones que cada usuario ha dado a un ítem i . Además, permite elegir diferentes métricas de similitud y métodos de predicción:

Métricas de similitud: coeficiente de Pearson, distancia coseno, distancia euclídea.

Métodos de predicción: simple, diferencia con la media.

Para acceder al sistema de forma online visite el siguiente [enlace](#)

Formato de la matriz

Las matrices de utilidad que usa el sistema deben tener el siguiente formato:

```
5 3 4 4 -
3 1 2 3 3
4 5 4 3 5
3 - 1 5 4
1 5 5 2 1
3 - 2 3 3
5 3 4 4 -
3 3 - 5 4
4 - - - 5
1 5 5 2 1
```

Donde cada fila u es un usuario y cada columna i es una valoración del usuario u sobre el ítem i ($r(u, i)$). Los ítems para los que no existe una valoración se representan con un guión (-).

Para cargar una matriz en el sistema puede seleccionar un fichero de texto desde su árbol de directorios local o puede usar una de las matrices de ejemplo que se encuentran en la página:



Sistema Recomendador

Matriz de utilidad

Seleccionar archivo

Ningún archivo seleccionado

Usar un ejemplo

Ejemplo 1.
Matriz de
utilidad 5x5.
1 incógnita

Ejemplo 2.
Matriz de
utilidad 5x5.
1 incógnita

Ejemplo 3.
Matriz de
utilidad 10x5.
8 incógnitas

A continuación puede elegir la métrica, método de predicción y número de vecinos a considerar:

Métrica

Correlación de Pearson



Número de vecinos considerados

3

Método de predicción

Diferencia con la media





Matriz de similitud

Una vez se ha cargado la matriz de utilidad, se habilitarán las opciones de visualizar esta y la matriz de similitud entre usuarios al fondo de la página:

Calcular

[Matriz de utilidad](#)

#	Item1	Item2	Item3	Item4	Item5
User1	5.00	3.00	4.00	4.00	-
User2	3.00	1.00	2.00	3.00	3.00
User3	4.00	3.00	4.00	3.00	5.00
User4	3.00	3.00	1.00	5.00	4.00
User5	1.00	5.00	5.00	2.00	1.00

[Matriz de similitud](#)

#	User1	User2	User3	User4	User5
User1	1.00	0.84	0.61	0.00	-0.77
User2	0.84	1.00	0.47	0.49	-0.90
User3	0.61	0.47	1.00	-0.16	-0.47
User4	0.00	0.49	-0.16	1.00	-0.64
User5	-0.77	-0.90	-0.47	-0.64	1.00



En este caso, al usar como métrica de similitud el coeficiente de correlación de Pearson, la diagonal consta de unos (correlación directa perfecta) como es lógico. Además, se puede apreciar la simetría de la matriz, lo que indica que no sería necesario calcularla completamente.

En el caso de usar la distancia euclídea como métrica de similitud se puede observar como la diagonal consta de ceros (correspondientes a un mismo punto):

[Matriz de similitud](#)

#	User1	User2	User3	User4	User5
User1	0.00	3.61	1.41	3.74	5.00
User2	3.61	0.00	3.61	3.16	5.83
User3	1.41	3.61	0.00	3.87	5.57
User4	3.74	3.16	3.87	0.00	6.48
User5	5.00	5.83	5.57	6.48	0.00

Selección de vecinos

En el apartado *Vecinos ordenados* puede visualizar los vecinos ordenados para cada usuario desde el más parecido al que menos. Aquellos vecinos coloreados en azul son los k vecinos más parecidos al usuario en cuestión, y son los tenidos en cuenta para la predicción del ítem i siempre y cuando todos hayan valorado dicho ítem, en caso de que uno de los vecinos no

haya valorado el ítem se toma el próximo vecino más parecido (si lo hay), en otro caso, se toma un número de vecinos menor que k.

[Vecinos ordenados](#)

#	1	2	3	4
User1	[2]: 0.84	[3]: 0.61	[4]: 0.00	[5]: -0.77
User2	[1]: 0.84	[4]: 0.49	[3]: 0.47	[5]: -0.90
User3	[1]: 0.61	[2]: 0.47	[4]: -0.16	[5]: -0.47
User4	[2]: 0.49	[1]: 0.00	[3]: -0.16	[5]: -0.64
User5	[3]: -0.47	[4]: -0.64	[1]: -0.77	[2]: -0.90



El número que está entre corchetes representa el índice del vecino, así pues, en este ejemplo, el vecino más similar al *User1* sería el *User2*, con una similitud de 0.84.

Puede saber cuáles vecinos se han tenido en cuenta para el cálculo de la predicción de cada ítem viendo el log que se genera al final de la página cada vez que se calcula una predicción:

[Log](#)

Para calcular la valoración del usuario 1 sobre el ítem 5 se han tenido en cuenta 3 vecinos: 2,3,4 -> 4.85

Por ejemplo, en el caso de esta matriz de utilidad:

[Matriz de utilidad](#)

#	Item1	Item2	Item3	Item4	Item5	Item6	Item7	Item8	Item9	Item10
User1	0.00	0.00	2.00	3.00	4.00	4.00	5.00	3.00	0.00	0.00
User2	3.00	-	5.00	4.00	-	-	1.00	5.00	-	0.00
User3	5.00	-	2.00	4.00	3.00	-	0.00	-	-	4.00
User4	0.00	3.00	4.00	2.00	4.00	4.00	-	3.00	0.00	1.00
User5	-	-	5.00	4.00	5.00	4.00	3.00	4.00	2.00	5.00

Aunque el número de vecinos definido es 3, para el ítem 2 solo hay dos usuarios que lo han valorado, por lo que es imposible tener 3 vecinos, en este caso se hace el cálculo con los dos vecinos más parecidos:

[Log](#)

Para calcular la valoración del usuario 2 sobre el ítem 2 se han tenido en cuenta 2 vecinos: 4,1 -> 3.12

Para calcular la valoración del usuario 2 sobre el ítem 5 se han tenido en cuenta 3 vecinos: 4,1,5 -> 4.61

Para calcular la valoración del usuario 2 sobre el ítem 6 se han tenido en cuenta 3 vecinos: 4,1,5 -> 4.47

Para calcular la valoración del usuario 2 sobre el ítem 9 se han tenido en cuenta 3 vecinos: 4,1,5 -> 0.75

Para calcular la valoración del usuario 3 sobre el ítem 2 se han tenido en cuenta 2 vecinos: 1,4 -> 3.64

Para calcular la valoración del usuario 3 sobre el ítem 6 se han tenido en cuenta 3 vecinos: 5,1,4 -> 1.60

Para calcular la valoración del usuario 3 sobre el ítem 8 se han tenido en cuenta 3 vecinos: 5,2,1 -> 2.66

Para calcular la valoración del usuario 3 sobre el ítem 9 se han tenido en cuenta 3 vecinos: 5,1,4 -> 4.32

Para calcular la valoración del usuario 4 sobre el ítem 7 se han tenido en cuenta 3 vecinos: 1,5,2 -> 2.48

Para calcular la valoración del usuario 5 sobre el ítem 1 se han tenido en cuenta 3 vecinos: 4,3,2 -> 3.49

Para calcular la valoración del usuario 5 sobre el ítem 2 se han tenido en cuenta 2 vecinos: 4,1 -> 4.39



Predicciones

Para calcular las predicciones en los valores marcados como un guión al principio en una matriz de utilidad simplemente cargue la matriz y presione calcular, y se habilitará un apartado al final de la página donde visualizar las predicciones con fondo verde:

[Matriz de utilidad con predicciones](#)

#	Item1	Item2	Item3	Item4	Item5
User1	5.00	3.00	4.00	4.00	4.85
User2	3.00	1.00	2.00	3.00	3.00
User3	4.00	3.00	4.00	3.00	5.00
User4	3.00	3.00	1.00	5.00	4.00
User5	1.00	5.00	5.00	2.00	1.00

En este caso, se ha utilizado como métrica la correlación de Pearson, tres vecinos y se tiene en cuenta la diferencia con la media valoración de los usuarios.

Veamos como varía la predicción al modificar los parámetros:

- **Métrica:** Distancia coseno
- **Nº de vecinos:** 4
- **Método de predicción :** simple

#	Item1	Item2	Item3	Item4	Item5
User1	5.00	3.00	4.00	4.00	4.01
User2	3.00	1.00	2.00	3.00	3.00
User3	4.00	3.00	4.00	3.00	5.00
User4	3.00	3.00	1.00	5.00	4.00
User5	1.00	5.00	5.00	2.00	1.00



- **Métrica:** distancia euclídea
- **Nº de vecinos:** 4
- **Método de predicción:** simple

[Matriz de utilidad con predicciones](#)

#	Item1	Item2	Item3	Item4	Item5
User1	5.00	3.00	4.00	4.00	2.75
User2	3.00	1.00	2.00	3.00	3.00
User3	4.00	3.00	4.00	3.00	5.00
User4	3.00	3.00	1.00	5.00	4.00
User5	1.00	5.00	5.00	2.00	1.00

Como se puede observar, la predicción cambia drásticamente al utilizar la distancia euclídea como métrica de similitud, debido a que es una fórmula mucho menos sofisticada.

Descargar informe

Al calcular la matriz con los valores faltantes predecidos, se habilitará la opción de descargar los resultados en un informe en formato JSON que contiene los campos:

utilityMatrix,

similarityMatrix, metric,

predictionMethod,

numOfNeighbors y

predictedMatrix

Descargar

```
{
  "utilityMatrix":
  [
    [5,3,4,4,null],
    [3,1,2,3,3],
    [4,3,4,3,5],
    [3,3,1,5,4],
    [1,5,5,2,1]
  ],
  "similarityMatrix":
  [
    [1,0.8391813582966892,0.6063390625908326,0,-0.7680953726425684],
    [0.8391813582966892,1,0.46770717334674267,0.48995593493886586,-0.9001487972234684],
    [0.6063390625908326,0.46770717334674267,1,-0.161164592805076,-0.4665694748158435],
    [0.48995593493886586,-0.161164592805076,1,-0.6415029025857747],
    [-0.7680953726425684,-0.9001487972234684,-0.4665694748158435,-0.6415029025857747,1]
  ],
  "metric":"pearson",
  "predictionMethod":"userBased/collaborativeFiltering/meanDiff",
  "numOfNeighbors":3,
  "predictedMatrix":
  [
    [5,3,4,4,4.851676442821286],
    [3,1,2,3,3],
    [4,3,4,3,5],
    [3,3,1,5,4],
    [1,5,5,2,1]
  ]
}
```



Ejemplos

Matriz 5 x 10

Pearson, 3 vecinos, diferencia con la media.

[Matriz de similitud](#)

#	User1	User2	User3	User4	User5
User1	1.00	0.16	0.13	-0.21	0.65
User2	0.16	1.00	-0.69	-0.20	0.06
User3	0.13	-0.69	1.00	-0.34	0.27
User4	-0.21	-0.20	-0.34	1.00	-0.16
User5	0.65	0.06	0.27	-0.16	1.00

[Vecinos ordenados](#)

[Matriz de utilidad con predicciones](#)

#	Item1	Item2	Item3	Item4	Item5	Item6	Item7	Item8	Item9	Item10
User1	2.57	0.00	1.24	1.00	-1.13	1.00	4.00	0.00	2.00	0.00
User2	2.00	0.00	4.00	0.00	1.68	1.00	1.00	0.11	1.00	1.00
User3	2.00	3.00	0.00	2.00	1.00	3.00	2.00	0.00	4.00	3.00
User4	1.00	5.00	2.77	1.00	3.00	1.00	4.00	4.00	0.00	4.00
User5	5.00	2.56	3.00	2.95	0.00	1.00	5.00	3.00	5.00	3.00



Matriz 10 x 25

Coseno, 5 vecinos, simple

#	User1	User2	User3	User4	User5	User6	User7	User8	User9	User10
User1	1.00	0.70	0.68	0.66	0.71	0.63	0.68	0.50	0.65	0.77
User2	0.70	1.00	0.61	0.56	0.65	0.47	0.64	0.58	0.61	0.71
User3	0.68	0.61	1.00	0.63	0.56	0.60	0.68	0.73	0.56	0.72
User4	0.66	0.56	0.63	1.00	0.75	0.72	0.73	0.64	0.64	0.82
User5	0.71	0.65	0.56	0.75	1.00	0.63	0.78	0.49	0.69	0.69
User6	0.63	0.47	0.60	0.72	0.63	1.00	0.77	0.69	0.72	0.65
User7	0.68	0.64	0.68	0.73	0.78	0.77	1.00	0.55	0.59	0.59
User8	0.50	0.58	0.73	0.64	0.49	0.69	0.55	1.00	0.58	0.68
User9	0.65	0.61	0.56	0.64	0.69	0.72	0.59	0.58	1.00	0.70
User10	0.77	0.71	0.72	0.82	0.69	0.65	0.59	0.68	0.70	1.00

#	Item1	Item2	Item3	Item4	Item5	Item6	Item7	Item8	Item9	Item10	Item11	Item12	Item13	Item14	Item15	Item16	Item17	Item18	Item19	Item20	Item21	Item22	Item23	Item24	Item25
User1	2.00	5.00	5.00	4.00	0.00	2.00	0.00	5.00	0.00	1.00	2.00	4.00	4.00	0.00	2.00	4.00	3.00	5.00	0.00	4.00	1.00	4.00	0.00	3.00	5.00
User2	3.00	5.00	3.43	2.00	2.00	3.00	1.88	5.00	0.00	2.00	4.00	3.00	0.00	4.00	0.00	1.00	2.16	2.45	5.00	1.00	1.00	0.00	0.00	2.00	5.00
User3	1.00	4.00	4.00	1.00	0.00	3.63	5.00	1.00	3.00	1.00	1.00	2.00	5.00	4.00	1.00	0.00	4.00	5.00	1.00	3.00	4.00	0.00	1.00	1.00	2.00
User4	3.00	3.00	4.00	0.00	2.00	3.00	0.00	1.00	2.00	2.00	0.00	3.00	4.00	2.00	1.00	4.00	2.00	0.00	2.00	0.00	3.00	3.00	4.00	3.00	0.00
User5	2.00	0.00	5.00	0.00	3.00	1.00	0.00	4.00	3.00	5.00	5.00	5.00	4.00	3.00	1.00	2.00	1.00	2.00	1.00	0.00	1.00	5.00	1.00	5.00	3.00
User6	4.00	0.00	5.00	4.00	1.00	4.00	2.00	1.00	2.40	0.00	1.00	0.00	4.00	4.00	5.00	1.00	0.00	2.00	1.00	1.00	1.00	5.00	5.00	2.00	2.00
User7	5.00	2.29	4.00	4.00	4.00	4.00	1.36	3.00	2.00	1.84	5.00	0.00	5.00	5.00	1.00	1.00	5.00	1.00	0.00	2.00	1.00	4.00	1.00	4.00	0.00
User8	1.00	1.00	3.00	3.00	0.80	5.00	5.00	0.00	5.00	1.00	0.00	4.00	3.31	3.00	5.00	2.00	3.00	1.00	5.00	1.04	1.00	0.00	2.00	0.00	2.00
User9	1.00	1.00	1.00	2.00	0.00	2.82	0.00	5.00	2.00	2.00	1.00	2.00	0.00	3.00	3.00	0.00	2.00	2.00	1.00	0.98	3.00	4.00	5.00	3.02	3.00
User10	0.00	4.00	2.00	1.00	1.00	4.00	0.97	2.00	1.00	2.00	1.56	2.00	3.00	1.00	1.00	3.00	0.00	2.00	1.00	0.00	2.00	1.00	3.00	2.79	3.00

Coseno, 5 vecinos, diferencia con la media

#	Item1	Item2	Item3	Item4	Item5	Item6	Item7	Item8	Item9	Item10	Item11	Item12	Item13	Item14	Item15	Item16	Item17	Item18	Item19	Item20	Item21	Item22	Item23	Item24	Item25
User1	2.00	5.00	5.00	4.00	0.00	2.00	0.00	5.00	0.00	1.00	2.00	4.00	4.00	0.00	2.00	4.00	3.00	5.00	0.00	4.00	1.00	4.00	0.00	3.00	5.00
User2	3.00	5.00	3.40	2.00	2.00	3.00	1.83	5.00	0.00	2.00	4.00	3.00	0.00	4.00	0.00	1.00	2.14	2.42	5.00	1.00	1.00	0.00	0.00	2.00	5.00
User3	1.00	4.00	4.00	1.00	0.00	3.57	5.00	1.00	3.00	1.00	1.00	2.00	5.00	4.00	1.00	0.00	4.00	5.00	1.00	3.00	4.00	0.00	1.00	1.00	2.00
User4	3.00	3.00	4.00	0.00	2.00	3.00	0.00	1.00	2.00	2.00	0.00	3.00	4.00	2.00	1.00	4.00	2.00	0.00	2.00	0.00	3.00	3.00	4.00	3.00	0.00
User5	2.00	0.00	5.00	0.00	3.00	1.00	0.00	4.00	3.00	5.00	5.00	5.00	4.00	3.00	1.00	2.00	1.00	2.00	1.00	0.00	1.00	5.00	1.00	5.00	3.00
User6	4.00	0.00	5.00	4.00	1.00	4.00	2.00	1.00	2.50	0.00	1.00	0.00	4.00	4.00	5.00	1.00	0.00	2.00	1.00	1.00	1.00	5.00	5.00	2.00	2.00
User7	5.00	2.73	4.00	4.00	4.00	4.00	1.80	3.00	2.00	2.28	5.00	0.00	5.00	5.00	1.00	1.00	5.00	1.00	0.00	2.00	1.00	4.00	1.00	4.00	0.00
User8	1.00	1.00	3.00	3.00	1.09	5.00	5.00	0.00	5.00	1.00	0.00	4.00	3.61	3.00	5.00	2.00	3.00	1.00	5.00	1.28	1.00	0.00	2.00	0.00	2.00
User9	1.00	1.00	1.00	2.00	0.00	2.54	0.00	5.00	2.00	2.00	1.00	2.00	0.00	3.00	3.00	0.00	2.00	2.00	1.00	0.70	3.00	4.00	5.00	2.63	3.00
User10	0.00	4.00	2.00	1.00	1.00	4.00	0.48	2.00	1.00	2.00	1.10	2.00	3.00	1.00	1.00	3.00	0.00	2.00	1.00	0.00	2.00	1.00	3.00	2.24	3.00



Como se puede observar, las predicciones difieren mucho cuando se tiene en cuenta la media de calificaciones de los usuarios, es decir, si un usuario es propenso a valorar los ítems con notas más altas o bajas.

Debido a que sería difícil visualizar cómodamente matrices de grandes dimensiones en el formato de este documento, en el siguiente [enlace](#) puede encontrar una colección de informes generados al aplicar predicciones con distintas matrices y parámetros.