



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

Gestión del Conocimiento en las Organizaciones:

Sistema de recomendación.

Jeff Pérez Frade
(alu0101038520@ull.edu.es)



Índice:

1. Análisis realizado.	2
2. ¿Cómo evaluar la similitud?.	2
2.1. Correlación de Pearson.	2
2.2. Distancia Coseno.	3
2.3. Distancia Euclídea.	4
3. Cálculo de predicciones.	5
3.1. Predicción Simple.	5
3.2. Predicción diferencia con la media.	6
4. Ejemplos.	6
4.1. Mensaje de Error.	6
4.2. Ejemplo del programa.	7
4.3. Ejemplo matriz de utilidad 5 x 10.	9
4.4. Ejemplo de matriz de utilidad 10 x 25.	12
5. Conclusión.	14



1. Análisis realizado.

Para este sistema de recomendación hemos utilizado el enfoque basado en usuarios, es decir, las calificaciones proporcionadas por usuarios afines son empleadas para predecir la de uno en particular.

La técnica básica consiste en, dado un usuario activo, *Usuario 1*, y un ítem *i* el cual todavía no ha sido valorado por este *Usuario 1*, es necesario encontrar un conjunto de vecinos cercanos (otros usuarios) a quienes les haya gustado los mismos ítems que al *Usuario 1* en el pasado y ya hayan valorado el ítem *i*.

A continuación se utilizarán esas valoraciones para predecir la valoración que nuestro *Usuario 1* podría darle al ítem *i*. Este procedimiento debe ser repetido para todos los usuarios los cuales tengan valoraciones sin completar.

2. ¿Cómo evaluar la similitud?.

Contamos con varias medidas de similitud para su cálculo, en este caso utilizaremos tres, **correlación de Pearson**, **distancia coseno** y **distancia Euclídea**.

Estas calculan un índice que puede utilizarse para medir el grado de relación de dos variables siempre y cuando ambas sean continuas y cualitativas.

2.1. Correlación de Pearson.

Fórmula:

$$sim(u, v) = \frac{\sum_{i \in S_{uv}} (r(u, i) - \bar{r}(u)) \cdot (r(v, i) - \bar{r}(v))}{\sqrt{\sum_{i \in S_{uv}} (r(u, i) - \bar{r}(u))^2} \sqrt{\sum_{i \in S_{uv}} (r(v, i) - \bar{r}(v))^2}}$$

- ⇒ **u, v** son los usuarios a estudiar.
- ⇒ **r(u, i)** la calificación del usuario u del ítem i
- ⇒ **r(u)** la media de las calificaciones del usuario u
- ⇒ **S_{u,v}** conjunto de ítems calificados por **u** y **v**.



Código en JavaScript:

```
// Devuelve el valor de la correlacion de pearson entre dos usuarios
function correlacionPearson(u, v){
    // Medias de los usuarios
    let rU = mediaConjunta(u,v)[0];
    let rV = mediaConjunta(u,v)[1];

    // Numerador y denominadores de cada usuario
    let num = denU = denV = 0;
    for(var i = 0; i < u.length; i++){
        if(u[i] != '-' && v[i] != '-'){
            num += (u[i] - rU) * (v[i] - rV);
            denU += Math.pow((u[i] - rU),2);
            denV += Math.pow((v[i] - rV),2);
        }
    }
    let Pearson = num / (Math.sqrt(denU) * Math.sqrt(denV));
    return Pearson;
}
```

- ⇒ **u** y **v** son los usuarios a estudiar.
- ⇒ **Pearson** es el resultado de esa correlación.

Los valores posibles de similitud van desde -1 hasta 1:

Si **sim(u, v) = 1**, correlación directa perfecta.

Si **0 < sim(u, v) < 1**, correlación directa.

Si **sim(u, v) = 0**, no hay correlación.

Si **-1 < sim(u, v) < 0**, correlación inversa.

Si **sim(u, v) = -1**, correlación inversa perfecta.

2.2. Distancia Coseno.

Si dos vectores tienen exactamente la misma orientación, su coseno toma el valor de 1, si son perpendiculares su coseno es 0, y si tienen orientaciones opuestas su coseno es de -1.

Fórmula:

$$sim(u, v) = \frac{\sum_{i \in S_{uv}} r(u, i) \cdot r(v, i)}{\sqrt{\sum_{i \in S_{uv}} (r(u, i))^2} \sqrt{\sum_{i \in S_{uv}} (r(v, i))^2}}$$



Código en JavaScript:

```
// Devuelve el valor de similitud mediante la Distancia Coseno entre dos usuarios
function distanciaCoseno(u, v){
    let num = denU = denV = 0;

    for(var i = 0; i < u.length; i++){
        if(u[i] != '-' && v[i] != '-'){
            num += u[i] * v[i];
            denU += Math.pow(u[i],2);
            denV += Math.pow(v[i],2);
        }
    }
    let Coseno = num / (Math.sqrt(denU) * Math.sqrt(denV));
    return Coseno;
}
```

- ⇒ **u** y **v** usuarios a estudiar similitud mediante distancia coseno.
- ⇒ **Coseno** devuelve el valor de similitud entre ambos usuarios.

2.3. Distancia Euclídea.

Entre dos puntos p y q se define como la longitud del segmento que une ambos puntos. Puede generalizarse para un espacio Euclídeo n-dimensional.

Fórmula:

$$d(u, v)_{euc} = \sqrt{\sum_{p \in P} (r(u, i) - r(v, i))^2}$$

- ⇒ **P** conjunto de ítems clasificados por **u** y **v**

Código en JavaScript:

```
// Devuelve el valor de similitud mediante la Distancia Euclidea entre dos usuarios
function distanciaEuclidea(u, v){
    let sum = 0;
    for(var i = 0; i < u.length; i++){
        if(u[i] != '-' && v[i] != '-'){
            sum += Math.pow(u[i] - v[i],2);
        }
    }
    let Euclidea = Math.sqrt(sum);
    return Euclidea;
}
```



3. Cálculo de predicciones.

Es necesario calcular el valor desconocido de $\hat{r}(u, i)$ utilizando las puntuaciones asignadas a los ítems i de los usuarios v más parecidos, es decir, los más próximos.

3.1. Predicción Simple.

Hay que tener en cuenta que la predicción con simples promedios no tiene en cuenta las desviaciones.

Fórmula:

$$\hat{r}(u, i) = \frac{\sum_{v \in N_u^k} sim(u, v) \cdot r(v, i)}{\sum_{v \in N_u^k} |sim(u, v)|}$$

⇒ **N** representa el conjunto de **k** vecinos más próximos en términos de función de similitud $sim(u, v)$ ya sea por Pearson, Coseno o Euclídea. Por lo general necesitamos un número mínimo de **k** vecinos, al menos 3.

Código en JavaScript:

```
for(var i = 0; i < nVecinos; i++){
    let vecinoActual = vecinosProx[i].vecino;
    if(matriz[vecinoActual][item] != '-') { // Compruebo si el vecino tiene la valoracion del Item
        num += vecinosProx[i].sim * matriz[vecinoActual][item] ;
        den += Math.abs(vecinosProx[i].sim);
        elecciones.innerHTML += `<p>Cálculo del usuario ${u} con el vecino mas próximo ${vecinoActual} </p>`;
    } else if(nVecinos != vecinosProx.length) nVecinos++;
}
elecciones.innerHTML += `<p>Predicción: ${num/den}</p>`;
return num / den;
```

⇒ **vecinosProx[]** es un objeto que contiene los vecinos más próximos ordenados por similitud de cada usuario que se estudia.



3.2. Predicción diferencia con la media.

Este tipo de predicción si tiene en cuenta las desviaciones.

Fórmula:

$$\hat{r}(u, i) = \bar{r}(u) + \frac{\sum_{v \in N_u^k} sim(u, v) \cdot (r(v, i) - \bar{r}(v))}{\sum_{v \in N_u^k} |sim(u, v)|}$$

⇒ **r(u)** representa la media de puntuaciones.

Incluir la media del usuario activo y del vecindario es una solución para compensar las diferencias de interpretación.

Código en JavaScript:

```
for(var i = 0; i < nVecinos; i++){
    let vecinoActual = vecinosProx[i].vecino;
    if(matriz[vecinoActual][item] != '-') { // Compruebo si el vecino tiene la valoracion del Item
        num += vecinosProx[i].sim * (matriz[vecinoActual][item] - media(matriz[vecinoActual]));
        den += Math.abs(vecinosProx[i].sim);
        elecciones.innerHTML += `<p>Cálculo del usuario ${u} con el vecino mas próximo ${vecinoActual} </p>`;
    } else if(nVecinos != vecinosProx.length) nVecinos++;
}
elecciones.innerHTML += `<p>Predicción: ${media(matriz[u]) + (num / den)}</p>`;
return media(matriz[u]) + (num / den);
```

⇒ **vecinosProx[]** es un objeto que contiene los vecinos más próximos ordenados por similitud de cada usuario que se estudia.

4. Ejemplos.

Aquí proporcionaremos algunos ejemplos intercambiando todos los aspectos de la aplicación.

4.1. Mensaje de Error.

La aplicación está diseñada para lanzar un mensaje de error si se introduce un número de vecinos superior al número de usuarios a estudiar.

Aquí vemos como intento introducir un número de vecinos a estudiar de 10 en una matriz de **10 x 25**.



Sistema de Recomendación

Seleccionar archivo

Selecciona una métrica:

Selecciona el tipo de predicción:

Escribe un número

Matriz de Utilidad

3	5	0	1	4	5	5
4	5	1	3	4	2	0
2	5	1	1	1	3	2
3	5	1	4	1	2	3
5	1	0	3	0	0	5
2	4	3	3	0	0	0
1	5	1	5	1	5	1
4	1	1	3	-	2	1
1	5	0	1	1	0	3
1	0	2	5	1	5	1

Error
El número de vecinos a estudiar debe ser inferior al número de usuarios

4.2. Ejemplo del programa.

Primero hacemos una prueba con la matriz de utilidad que se encuentra en las transparencias de clase. Hemos elegido el archivo que contiene la matriz, seleccionado el tipo de métrica, el tipo de predicción y el número de vecinos.

Métrica: Correlación de Pearson.

Predicción: Diferencia con la media.

Número de vecinos: 2.



Sistema de Recomendación

Seleccionar archivo original.txt

Selecciona una métrica:

Correlación de Pearson

Selecciona el tipo de predicción:

Diferencia con la media

Escribe un número de vecinos:

2

Matriz de Utilidad cargada:

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

Ejecutar

Matriz resultante:

```
5 3 4 4 4.87  
3 1 2 3 3  
4 3 4 3 5  
3 3 1 5 4  
1 5 5 2 1
```

Observamos que debajo de la matriz resultante tenemos las comparaciones con sus vecinos más próximos y su similitud.

Estudio de predicción para el Item 4 del Usuario 0

Cálculo del usuario 0 con el vecino mas próximo 1 con similitud 0.8528028654224417

Cálculo del usuario 0 con el vecino mas próximo 2 con similitud 0.7071067811865475

Predicción: 4.871979899370592



4.3. Ejemplo matriz de utilidad 5 x 10.

Métrica: Distancia Coseno.

Predicción: Simple.

Número de vecinos: 3.

Sistema de Recomendación

Seleccionar archivo m1.txt

Selecciona una métrica:

Distancia Coseno

Selecciona el tipo de predicción:

Predicción Simple

Escribe un número de vecinos:

3

Matriz de Utilidad cargada:

```
4 - - 0 2 - 3 - 0 -  
- 4 4 1 1 3 0 - - 2  
2 5 1 2 1 5 5 5 2 0  
1 4 1 3 1 - 1 0 0 0  
0 3 4 0 0 5 5 4 5 -
```

Ejecutar

Matriz resultante:

4	4.22	2.02	0	2	4.29	3	3.12	0	0.69
1.74	4	4	1	1	3	0	2.39	2.27	2
2	5	1	2	1	5	5	5	2	0
1	4	1	3	1	4.11	1	0	0	0
0	3	4	0	0	5	5	4	5	0.87



Similaridad entre cada uno de los usuarios junto con la predicción final:

Estudio de prediccion para el Item 1 del Usuario 0

Cálculo del usuario 0 con el vecino mas próximo 2 con similitud 0.7530940414109725

Cálculo del usuario 0 con el vecino mas próximo 3 con similitud 0.4824506406770077

Cálculo del usuario 0 con el vecino mas próximo 4 con similitud 0.39391929857916763

Predicción: 4.220425088920805

Estudio de prediccion para el Item 2 del Usuario 0

Cálculo del usuario 0 con el vecino mas próximo 2 con similitud 0.8489238841769257

Cálculo del usuario 0 con el vecino mas próximo 1 con similitud 0.8017355944006131

Cálculo del usuario 0 con el vecino mas próximo 3 con similitud 0.714863886849203

Predicción: 2.0167757454249196

Estudio de prediccion para el Item 5 del Usuario 0

Cálculo del usuario 0 con el vecino mas próximo 2 con similitud 0.8431882126488588

Cálculo del usuario 0 con el vecino mas próximo 1 con similitud 0.7827758406576656

Cálculo del usuario 0 con el vecino mas próximo 4 con similitud 0.5785125855697117

Predicción: 4.28983067740133

Estudio de prediccion para el Item 7 del Usuario 0

Cálculo del usuario 0 con el vecino mas próximo 2 con similitud 0.8858957766689782

Cálculo del usuario 0 con el vecino mas próximo 3 con similitud 0.7262633011177178

Cálculo del usuario 0 con el vecino mas próximo 4 con similitud 0.6870300416352265

Predicción: 3.121795849351635

Estudio de prediccion para el Item 9 del Usuario 0

Cálculo del usuario 0 con el vecino mas próximo 2 con similitud 0.8973164911239071

Cálculo del usuario 0 con el vecino mas próximo 1 con similitud 0.8320346289648878

Cálculo del usuario 0 con el vecino mas próximo 3 con similitud 0.6654045647822469

Predicción: 0.6948805961470705



Estudio de prediccion para el Item 0 del Usuario 1

Cálculo del usuario 1 con el vecino mas próximo 0 con similitud 0.8197618547317124

Cálculo del usuario 1 con el vecino mas próximo 4 con similitud 0.7571877794400365

Cálculo del usuario 1 con el vecino mas próximo 3 con similitud 0.735767207381959

Predicción: 1.7359732735857165

Estudio de prediccion para el Item 7 del Usuario 1

Cálculo del usuario 1 con el vecino mas próximo 0 con similitud 0.815343405174656

Cálculo del usuario 1 con el vecino mas próximo 3 con similitud 0.7462279083203267

Cálculo del usuario 1 con el vecino mas próximo 4 con similitud 0.7318610910490817

Predicción: 2.3856451044734217

Estudio de prediccion para el Item 8 del Usuario 1

Cálculo del usuario 1 con el vecino mas próximo 0 con similitud 0.8356201673143685

Cálculo del usuario 1 con el vecino mas próximo 4 con similitud 0.7659893074039367

Cálculo del usuario 1 con el vecino mas próximo 2 con similitud 0.7334325162818318

Predicción: 2.268400992358444

Estudio de prediccion para el Item 5 del Usuario 3

Cálculo del usuario 3 con el vecino mas próximo 2 con similitud 0.6889283267815868

Cálculo del usuario 3 con el vecino mas próximo 1 con similitud 0.6635242933863315

Cálculo del usuario 3 con el vecino mas próximo 0 con similitud 0.6628070008766147

Predicción: 4.107985125775844

Estudio de prediccion para el Item 9 del Usuario 4

Cálculo del usuario 4 con el vecino mas próximo 2 con similitud 0.8609021141167816

Cálculo del usuario 4 con el vecino mas próximo 1 con similitud 0.7867050394437695

Cálculo del usuario 4 con el vecino mas próximo 0 con similitud 0.7276595671437707

Predicción: 0.8737945773101675



4.4. Ejemplo de matriz de utilidad 10 x 25.

Métrica: Distancia Euclídea.

Predicción: Diferencia con la media.

Número de vecinos: 7.

Selecciona una métrica:

Distancia Euclídea

Selecciona el tipo de predicción:

Diferencia con la media

Escribe un número de vecinos:

7

Matriz de Utilidad cargada:

```
3 5 0 1 4 5 5 1 2 4 0 0 1 4 1 3 2 4 1 1 5 - 5 2 0
4 5 1 3 4 2 0 1 4 - 0 2 - 2 0 1 4 5 3 - 1 0 5 - 0
2 5 1 1 1 3 2 4 5 3 4 1 1 3 - 1 5 - 5 5 4 1 1 5 5
3 5 1 4 1 2 3 - 5 5 2 5 1 5 3 2 3 3 1 3 1 5 5 - 3
5 1 0 3 0 0 5 0 4 2 3 1 5 4 5 4 2 5 3 3 - 3 3 5 0
2 4 3 3 0 0 0 2 0 5 2 0 4 2 1 5 3 4 3 1 3 2 5 0 2
1 5 1 5 1 5 1 - 5 5 0 2 1 1 - 5 0 3 1 3 - 2 5 2 3
4 1 1 3 - 2 1 1 5 2 0 1 1 5 0 0 4 5 2 3 0 - 1 3 5
1 5 0 1 1 0 3 2 0 5 2 1 2 - 3 3 5 1 2 - 5 3 5 3 1
1 0 2 5 1 5 1 2 2 2 2 3 4 5 2 5 3 3 3 0 3 3 4 4 2
```

Ejecutar

Matriz resultante:

3	5	0	1	4	5	5	1	2	4	0	0	1	4	1	3	2	4	1	1	5	2.11	5	2	0
4	5	1	3	4	2	0	1	4	3.19	0	2	2.24	2	0	1	4	5	3	1.89	1	0	5	2.8	0
2	5	1	1	1	3	2	4	5	3	4	1	1	3	2.53	1	5	4.23	5	5	4	1	1	5	5
3	5	1	4	1	2	3	2.27	5	5	2	5	1	5	3	2	3	3	1	3	1	5	5	3.61	3
5	1	0	3	0	0	5	0	4	2	3	1	5	4	5	4	2	5	3	3	2.98	3	3	5	0
2	4	3	3	0	0	0	2	0	5	2	0	4	2	1	5	3	4	3	1	3	2	5	0	2
1	5	1	5	1	5	1	1.78	5	5	0	2	1	1	2.08	5	0	3	1	3	2.75	2	5	2	3
4	1	1	3	0.77	2	1	1	5	2	0	1	1	5	0	0	4	5	2	3	0	1.89	1	3	5
1	5	0	1	1	0	3	2	0	5	2	1	2	3.32	3	3	5	1	2	2.48	5	3	5	3	1
1	0	2	5	1	5	1	2	2	2	2	3	4	5	2	5	3	3	3	0	3	3	4	4	2



Similitud con los vecinos (**solo una captura debido a su extensión**):

Estudio de prediccion para el Item 21 del Usuario 0

Cálculo del usuario 0 con el vecino mas próximo 2 con similitud 12.529964086141668

Cálculo del usuario 0 con el vecino mas próximo 4 con similitud 11.789826122551595

Cálculo del usuario 0 con el vecino mas próximo 9 con similitud 11.269427669584644

Cálculo del usuario 0 con el vecino mas próximo 5 con similitud 11.180339887498949

Cálculo del usuario 0 con el vecino mas próximo 3 con similitud 10.392304845413264

Cálculo del usuario 0 con el vecino mas próximo 6 con similitud 9.539392014169456

Cálculo del usuario 0 con el vecino mas próximo 1 con similitud 9.055385138137417

Predicción: 2.1063172960923366

Estudio de prediccion para el Item 9 del Usuario 1

Cálculo del usuario 1 con el vecino mas próximo 4 con similitud 11.357816691600547

Cálculo del usuario 1 con el vecino mas próximo 9 con similitud 10.954451150103322

Cálculo del usuario 1 con el vecino mas próximo 8 con similitud 10.723805294763608

Cálculo del usuario 1 con el vecino mas próximo 2 con similitud 10.295630140987

Cálculo del usuario 1 con el vecino mas próximo 3 con similitud 9.797958971132712

Cálculo del usuario 1 con el vecino mas próximo 6 con similitud 9.327379053088816

Cálculo del usuario 1 con el vecino mas próximo 0 con similitud 9.297962142319143

Predicción: 3.19412583738597

Estudio de prediccion para el Item 12 del Usuario 1

Cálculo del usuario 1 con el vecino mas próximo 4 con similitud 11.419986865141308

Cálculo del usuario 1 con el vecino mas próximo 9 con similitud 11.01889740400554

Cálculo del usuario 1 con el vecino mas próximo 8 con similitud 10.875481598531625

Cálculo del usuario 1 con el vecino mas próximo 2 con similitud 10.297383162726344

Cálculo del usuario 1 con el vecino mas próximo 3 con similitud 9.963739257929223

Cálculo del usuario 1 con el vecino mas próximo 6 con similitud 9.501373584908658

Cálculo del usuario 1 con el vecino mas próximo 5 con similitud 9.395536174162707

Predicción: 2.2390925944777655



5. Conclusión.

Los sistemas de recomendación son herramientas de gran utilidad a la hora de ayudar a los usuarios en sus procesos de búsqueda. Después de estudiar los dos grupos de sistemas, los colaborativos y los no colaborativos, se llega a la conclusión de que en muchas ocasiones la mejor opción es una clase de enfoque híbrido para aprovechar las ventajas de ambos.

El desarrollo de estos sistemas será muy importante en el futuro ya que diariamente se incorporan más usuarios a la web, sumado a la creación de más sitios de comercio electrónico hace que contar con estos sistemas sea clave para la búsqueda de información y la toma de decisiones.