similitudes-peliculas.py

Tenemos un fichero con opiniones de usuarios acerca de películas. El formato es:

input

```
196 242 3   881250949
186 302 3   891717742
22  377 1   878887116
244 51  2   880606923
    …
```

Los campos son: 1) id de usuario, 2) id de película, 3) puntuación, 4) timestamp.

Por cada par de películas, queremos establecer una relación entre ellas para indicar si son similares o no. Con esto, cuando nos muestren una película podemos mostrar aquellas que son similares.

En esta primera fase, vamos a partir de una película y ver la distancia que hay a otra.

El procedimiento para obtener la similitud es

1. Pasar de
   a. *id usuario*, *id película*, *puntuación*, *timestamp* a
   b. *id usuario* => *(id película, puntuación)*
2. Producto cartesiano. Por cada par de películas de un usuario tenemos ahora:
   (id usuario, ((id película1, puntuación1), (id película2, puntuación2)))
3. Filtramos los duplicados. Cada película está asociada a sí misma y por cada par de películas hay dos entradas. Filtramos y elegimos las entradas en las que *id película1* sea menor que *id película2*.
4. Hacemos un mapeo para que las claves sean los pares de películas; pasamos de
   id usuario => ((id película1, puntuación1), (id película2, puntuación2)) a
   (id pelicula1, id pelicula2) => (puntuación1, puntuación2)
   Es decir, nos olvidamos ya del usuario.
5. Agrupamos los pares de películas. Es decir el par de películas es la clave y el valor todos los pares de opiniones sobres las películas:
   (pelicula1, pelicula2) => (opinion1, opinion2), (opinion1, opinion2)
6. Ahora ejecutamos la similitud coseno para averiguar los valores. El resultado es un par (valor, número de veces) en el que valor es realmente el resultado y número de veces son las veces que ese par ha sido calificado por un usuario. Por ejemplo, si número de veces es 1 indica que sólo un usuario ha calificado ese par de películas con lo que no se debería tomar en cuenta la similitud.
7. Al final mostramos las películas similares ordenadas por precisión.