Laboratori de CAIM

Pràctica 6

Walter J. Troiani

Prof: Ignasi Gómez Sebastià

2/12/2023 - 2023/24 Q1

1 Implementació d'un sistema de recomanació

L'objectiu d'aquesta pràctica i experimentació, serà construir un sistema de recomanació de pel·lícules, basant-nos en l'algorisme de filtratge col·laboratiu de tipus "user-to-user". Aquest es basa en les recomanacions d'items desconeguts mitjançant les puntuacions que han donat els "veïns" o usuaris més similars del sistema. En les nostres dades les puntuacions dels usuaris venen donades del 0 al 5 (Una anàlisi més exhaustiu pot ser trobat a l'apartat 2.1). La formula per saber predir l'interès d'un usuari per un cert objecte es la seguent per un usuari u i un objecte i qualssevol:

$$Interest(u,i) = \bar{r}_u + \frac{\sum_{v \in N(u,i)} \text{Similarity}(u,v) \cdot (r_{v,i} - \bar{r}_v)}{\sum_{v \in N(u,i)} \text{Similarity}(u,v)}$$

Les recomanacions es faran d'acord amb les pel·lícules que l'usuari desitjat no ha vist i tinguent en compte les puntuacions donades pels seus veïns. Aleshores per garantir una certa qualitat d'aquestes recomanacions serà crític escollir una mesura de similitud bona i escollir un bon nombre de pel·lícules a recomanar i sobretot un nombre adequat de quants veïns hauria de tenir un usuari. Com no tenim accés al feedback d'usuaris l'experimentació s'haurà de basar en criteris propis i qualitatius que no depenguin dels feedback, igual que en la pràctica de Rocchio.

2 Experimentació

En aquest apartat ens centrarem a experimentar amb els possibles paràmetres rellevants

- ϕ, la mida del veïnat (Neighborhood) d'usuaris més semblants a considerar d'un usuari concret.
- μ, la mida del conjunt de recomanacions (Ítems més rellevants primer)
- θ, el nombre de gèneres més importants per a cada usuari a tenir en compte
- **sim**, el criteri de similitud entre usuaris emprat

Cal recalcar que aquesta experimentació es farà fixant els paràmetres i fixant els documents d'entrada al conjunt de dades de pel·lícules proporcionats per MovieLens. Per estalviar recursos en l'experimentació, la qual pot arribar a ser computacionalment costosa, farem servir la versió reduïda amb propòsit educacional. Els

paràmetres fixats per defecte han sigut fruit de l'experimentació i exploració manual de diverses execucions. La màquina on ha estat executat és un processador Intel i7-1265 amb 10 nuclis i 32 GB de RAM (Powered by Mango)

2.1 Anàlisi exploratòria de les dades

El primer pas per a construir el sistema de recomanació va ser familiaritzar-se amb les dades, tant en el format, la quantitat de nuls, la dispersió d'aquests (Sparsity) i altres mesures estadístiques com distribucions. Això és necessari, tant per la construcció del sistema, saber quines operacions són adients tant com per la posterior anàlisi dels resultats, com aquesta anàlisi i processat es llarg, el deixem fora de l'abast d'aquest escrit, però s'adjunta un jupyter notebook interactiu anomenat "explorative analysis.ipynb" on el lector es pot informar i llegir les conclusions extretes.

2.2 Anàlisi de les recomanacions

El factor més important a estudiar d'aquest sistema és sense cap mena de dubte la qualitat de les recomanacions, la qual pot ser més difícil de quantificar del que sembla. De manera qualitativa una persona pot jutjar si una recomanació és molt dolenta (O a primera vista ho sembla), però la resta de casos cauen en una zona grisa subjectiva. Per evitar al màxim el nostre biaix, partirem les dades en 2 conjunts: "train", on hi ha la majoria de les dades i "validation" on hi ha només 5 puntuacions (Per a cada usuari) escollits a l'atzar, per poder contrastar amb les pel·lícules recomanades. Però un altre criteri més important serà els gèneres que més mira l'usuari en comparació als gèneres dels films recomanats. Per l'experiment fixarem els paràmetres $\theta = \phi = \mu = 5, similitud = Pearson$

Com bé es pot veure, de totes les 25 pel·lícules recomanades només n'ha encertat una (Pero s'ha de tenir en compte que al tractar-se d'un mostreig uniforme la probabilitat es baixa, aleshores parlar de precisió no te molt sentit). Si podem veure que la comparació de les pel·lícules no es força positiva, ja que hi han usuaris com el 608 que mira pel·lícules d'un to més d'acció/bèl·lic i el sistema li recomana pel·lícules infantils i d'amor. Però si mirem també hi ha altres usuaris com el 32 que les prediccions tenen més sentit, però no és bon mètode per avaluar, a causa de l'aleatorietat, a més no sabem realment si és mala recomanació, simplement a simple vista ho sembla. Si mirem els gèneres, ens adonem que quasi sempre sol clavar els 3 primers i

Usuari	Gèneres Més Vists	Gèneres Recomanats
305	Drama: 301 Action: 282 Thriller: 248 Sci-Fi: 167 Comedy: 165	Drama: 4 Action: 1 War: 1 Thriller: 1 Animation: 1
64	Drama: 236 Comedy: 212 Thriller: 137 Action: 121 Romance: 106	Crime: 2 Thriller: 2 Comedy: 2 Drama: 2 Adventure: 2
32	Drama: 42 Comedy: 40 Thriller: 31 Romance: 30 Action: 29	Comedy: 3 Drama: 3 Crime: 2 Adventure: 1 Thriller: 1
595	Drama: 9 Comedy: 6 Romance: 5 Thriller: 3 Sci-Fi: 3	Drama: 4 Comedy: 4 Fantasy: 1
608	Comedy: 353 Drama: 278 Action: 276 Thriller: 258 Adventure: 181	Comedy: 5 Adventure: 3 Drama: 2 Romance: 2 Animation: 2

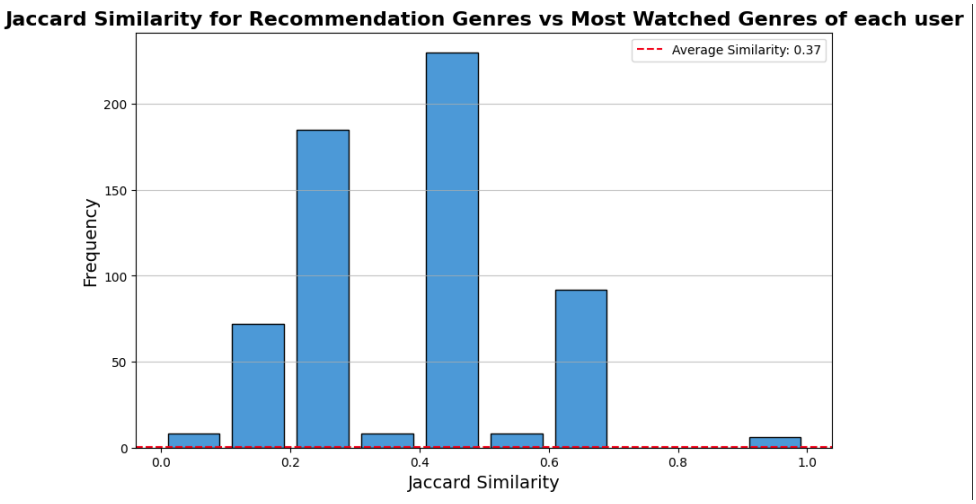
Table 1: Comparació de Gèneres Més Vists i Gèneres Recomanats per a diversos usuaris

Usuari	Pel·lícules Recomanades (Top 5)	Pel·lícules de Validació (Top 5)
305	Boot, Das (1981) Disclosure (1994) Ice Storm (1997) Simpsons Movie (2007) Hamlet (1996)	Look Who's Talking (1989) Critical Care (1997) Hamlet (1996) 20,000 Leagues Under the Sea (1954) Gods Must Be Crazy (1980)
64	Manchurian Candidate (1962) Entrapment (1999) Philadelphia Story (1940) 20,000 Leagues Under the Sea (1954) Gods Must Be Crazy (1980)	Look Who's Talking (1989) Critical Care (1997) Ponyo (2008) Smoke Signals (1998) NaN
32	Calendar Girl (1993) Stand by Me (1986) Lock, Stock & Two Smoking Barrels (1998) Midsummer Night's Dream (1999) Godfather (1972)	Swan Princess (1994) Robin Hood (1991) Gigi (1958) Aristocats (1970) City Slickers II
595	Sweet Hereafter (1997) Happiness (1998) Rushmore (1998) Election (1999) Being John Malkovich (1999)	Georgia (1995) Joe Gould's Secret (2000) 'night Mother (1986) Ready to Wear (1994) Artemisia (1997)
608	Ever After (1998) Sisterhood of the Traveling Pants (2005) 27 Dresses (2008) Madagascar (2005) Shrek the Third (2007)	Passion of Joan of Arc NaN Enemy of the State (1998) Banana Joe (1981) Full Metal Jacket (1987)

Table 2: Comparació de Pel·lícules Recomanades i Pel·lícules de Validació (Top 5) per a diversos usuaris

després els gèneres varien força més, pero de totes maneres això es millor que no pas la recomanació del recomanador naive.

Un cop vists uns quants exemples de recomanació, he volgut quantificar aquesta diferència entre els conjunts de gèneres, mitjançant la similitud de Jaccard que a pesar d'ignorar la importància de la freqüència de cada gènere, bàsicament ens permet saber com són de similars els conjunts. Aleshores hem aplicat aquesta similitud a totes les recomanacions de tots els usuaris del sistema i hem obtingut l'histograma següent (És adient, ja que es repeteixen molts valors):



Aquest gràfic exposa que hi ha ben poques prediccions perfectes o quasi (menys de 25), més de 300 per sobre de 0.4 i més de 200 amb similitud inferior a 0.3, donant un promig de 0.37, el qual es una similitud força petita, lo qual ens indica que hi ha poques prediccions perfectes, però no necessàriament es dolent, ja que si ens fixem en els exemples anteriors, sembla ser el cas que acerta els 3 més rellevants però després hi ha força varietat. Depenent de les opinions dels usuaris podria ser el cas que aquest tingui bones mètriques de **"novelty"** i "**serendipity**".

2.3 Anàlisi de la similitud entre usuaris

La similitud entre usuaris es pot fer de moltes maneres, però en aquest experiment considerarem la similitud de Jaccard, cosinus, cosinus ajustat i correlació de Pearson.

$$\text{Pearson}(u,v) = \frac{\sum_i (r_{u,i}-\bar{r}_u)(r_{v,i}-\bar{r}_v)}{\sqrt{\sum_i (r_{u,i}-\bar{r}_u)^2}\cdot\sqrt{\sum_i (r_{v,i}-\bar{r}_v)^2}}$$

$$\text{Adjusted Cosine}(u,v) = \frac{\sum_i (r_{u,i}-\bar{r}_i)(r_{v,i}-\bar{r}_i)}{\sqrt{\sum_i (r_{u,i}-\bar{r}_i)^2}\cdot\sqrt{\sum_i (r_{v,i}-\bar{r}_i)^2}}$$

Les dues primeres velles conegudes, però, ja des d'un punt de vista teòric el més probable és que ofereixin pobres resultats, l'exemple il·lustra com a pesar de ser usuaris molt similars en puntuació i generes, Jaccard dona una similitud petita, ja que el segon usuari ha puntuat poques pel·lícules i no li dona gens d'importància a. A més també veiem un exemple de per què la similitud cosinus pot ser mala idea (A causa de la seva poca sensibilitat de magnituds), donant una alta puntuació a usuaris molt diferents:

$$\mathbf{u} = [1,1,1,1,1], \mathbf{v} = [0,1,0,0,1]$$

$$\text{Jaccard}(\mathbf{u},\mathbf{v}) = \frac{\text{Intersecció}(\mathbf{u},\mathbf{v})}{\text{Unió}(\mathbf{u},\mathbf{v})} = \frac{1}{3}$$

$$\mathbf{u} = [5,5,5,5,5], \mathbf{v} = [1,1,1,1,1]$$

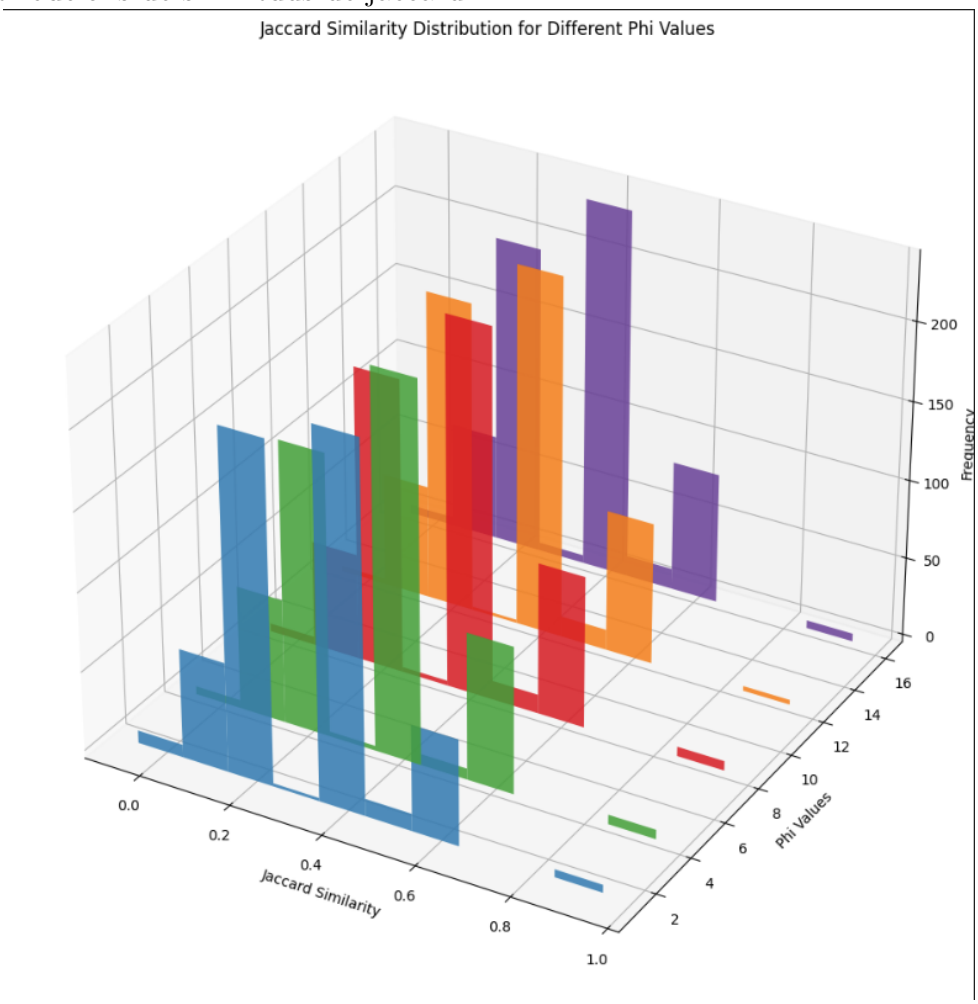
$$\text{Cosinus}(\mathbf{u},\mathbf{v}) = \frac{\mathbf{u}\cdot\mathbf{v}}{\|\mathbf{u}\|\cdot\|\mathbf{v}\|} = \frac{25}{\sqrt{125}\cdot\sqrt{5}} \approx 0.894$$

Evidentment, això ens deixa amb les dues distàncies restants que ambdues són força prometent, farem servir les dues en el script, on es pot apreciar les diferències en les similituds amb els seus veïns més semblants i també en les pel·lícules que recomana en conseqüència. Ambdues distàncies tenen un error força greu quan el nombre d'elements que comparteixen és 1 o petit, la similitud és total, cosa que no te gens de sentit, llavors probablement la millor mètrica seria aplicar una similitud cosinus o qualsevol altre si la intersecció és petita i Pearson/adjusted cosine si es suficient grossa.

2.4 Anàlisi de ϕ

Un factor força important en la qualitat de les recomanacions té a veure amb la mida del veïnat, per al qual hem experimentat amb diversos valors d'aquesta variable, fixant els paràmetres: $\theta = \mu = 5, similitud = Pearson$, obtenim les següents

distribucions de similituds de jaccard:

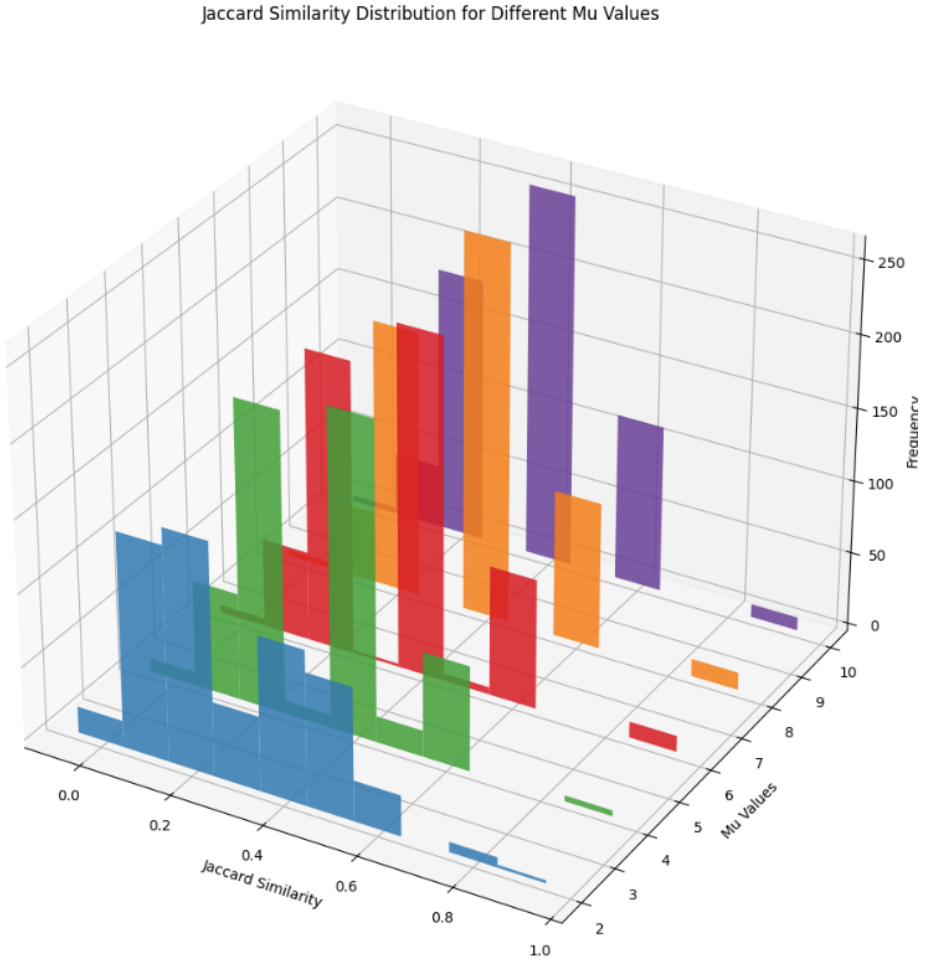


Això ens dona l'estranya conclusió que les distribucions són pràcticament iguals per contraintuïtiu que sembli. Una altra observació força evident és que contra més veïns més "wisdom of the crowd" s'aplicarà i pel·lícules més populars/estàndards es recomanaran, no tan personalitzades, faran una mena d'efecte mitjana.

2.5 Anàlisi de μ

Seguidament, analitzarem el pes del factor mu que té a veure amb la quantitat de pel·lícules (ordenades de més rellevància a menys) que el sistema recomanarà, ja que aquest influirà directament en la qualitat de la recomanació. Fixant els paràmetres: $\theta = \mu = 5, similitud = Pearson$, obtenim les següents distribucions de similituds de

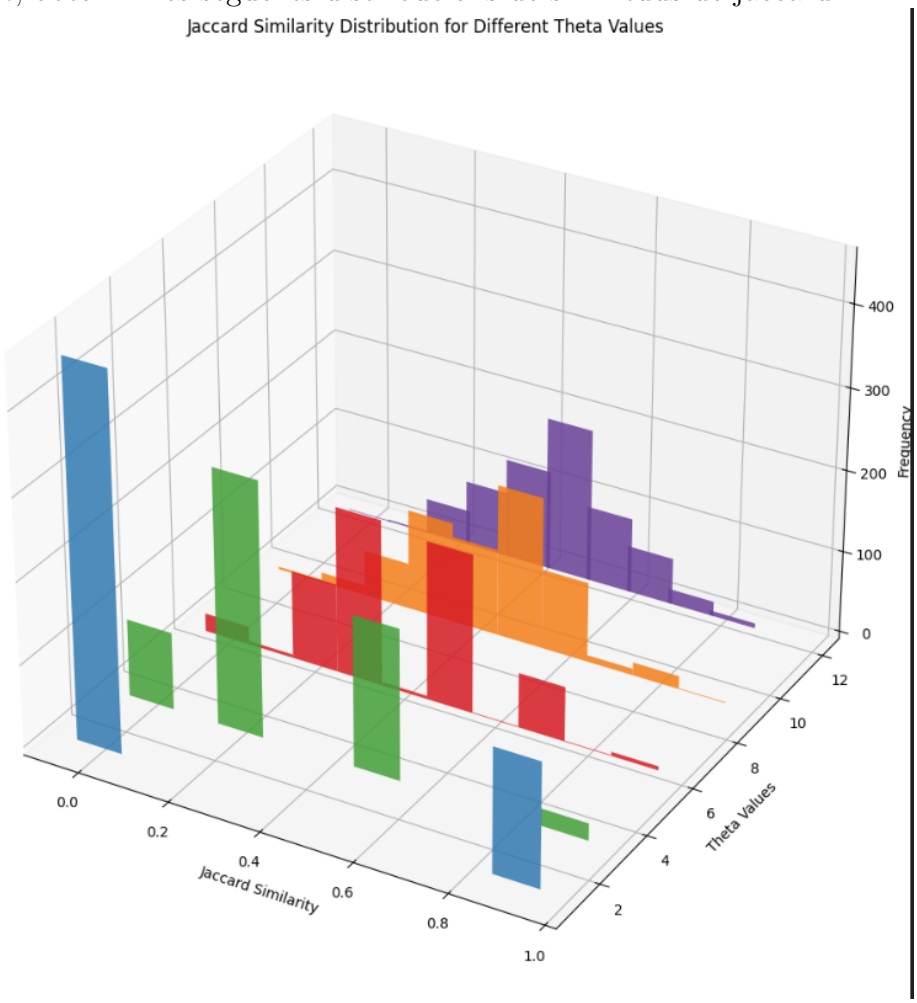
jaccard:



Observem que en valors baixos la similitud de jaccard mitjana és força més baixa que en els altres casos, això és degut a la quantitat de pel·lícules baixa fa que sigui més arriscada la predicció (Sobretot en el nostre sistema que veiem que té una tendència a recomanar diversitat i novetat). Però, en canvi, en valors més grans va augmentant (No massa després del primer augment), el qual té sentit, ja que contra més films s'hagin recomanat més generés hi haurà (Augmenta el recall).

2.6 Anàlisi de θ

Finalment, volem estudiar la importància del nombre de gèneres rellevants a considerar (θ, ja que aquests afectaran directament la nostra anàlisi (Però no té cap efecte significatiu en la recomanació). Fixant els paràmetres: $\phi = \mu = 5, similitud = Pearson$, obtenim les següents distribucions de similituds de jaccard:



Els resultats són força evidents i tenen sentit, contra menys generés hi ha més desparella es la distribució i contra més hi ha més se centren les dades en la mitjana, formant una mena de distribució normal.

3 Conclusions i dificultats

La dificultat més gran sense cap dubte ha estat l'avaluació del sistema de recomanació que no es gens fàcil quantificar com de bé està operant i menys encara sense possibilitat de feedback/ pseudofeedback (en forma de visualitzacions/interaccions/afegir a veure més tard..). També hagués sigut beneficiós emprar veïnats dinàmics o bé mètriques de "novelty", "diversity", "surprise", "serendipity" ... Pero la complexitat temporal de desenvolupar tot aquest codi i anàlisis creix força i donat que aquest equip d'investigadors només té un sol membre no hi ha hagut temps. Ni tan sols hi ha hagut temps a fer l'estudi de comparativa de Pearson Vs Adjusted Cosine o bé la mesura mixta com he mencionat a l'apartat anterior a causa d'aquesta manca de temps i possiblement haurien millorat la qualitat de les recomanacions.

També l'anàlisi exploratòria de les dades és concís i suficient, però si més recursos poguessin estar destinats, podria haver estat més exhaustiva. Un altre dificultat ha sigut desenvolupar el recomanador fins que funcionava realment bé, ja que quan hi havia errors en les transformacions de les dades i proporcionava resultats (Encara que fossin dolents o incomplets) era difícil trobar errors/bugs.

Com a conclusió final, és evident que el model és millor que el recomanador simple, pero podria ser encara millor si tinguéssim més dades, contra més dades i més qualitat millor. També podria ser millor si s'anés a primfilar als detalls i adaptar una mica el recomanador a com són les dades (Tal i com es fa en els sistemes reals) i no pas seguir l'esquema genèric del collaborative filtering. A més a més el esquema és senzill, evidentment un esquema híbrid emprant també content based filtering, K-Nearest Neighbors o algun esquema de xarxes neuronals.