

Disseny, implementació i estudi d'un sistema recomanador de pel·lícules

Arnau Ramon Moral

Escola Politècnica Superior d'Enginyeria de Vilanova i la Geltrú

Resum

En aquest Treball de Fi de Grau estudiarem la importància i rellevància dels sistemes recomanadors i veurem els diferents algoritmes que utilitzen avui dia les grans companyies tecnològiques. L'ús dels sistemes recomanadors s'ha vist molt incrementat per l'enorme quantitat de productes i/o serveis que poden arribar a oferir les grans empreses que operen online. La primera empresa que ens ve al cap, inevitablement, és Amazon, que ofereix més de 12 milions de productes per vendre. Sense un bon sistema recomanador de productes i d'anuncis, és fàcil que a un usuari se li recomanés justament el producte que vol comprar.

Tot i que els algoritmes recomanadors poden ser comuns i independents del que s'està recomanant (música, llibres, publicitat de productes, articles, etc.) en aquest treball ens centrarem en l'àmbit de les pel·lícules i les plataformes de vídeo mundialment conegudes (Netflix, HBO, Amazon Prime Video, Disney+, etc.).

A més de l'estudi de l'estat de l'art actual, dissenyarem i implementarem des de zero un sistema recomanador de pel·lícules que disposarà d'una petita interfície en la que es mostraran els resultats segons les preferències de l'usuari.

Per últim, realitzarem una anàlisi de les dades obtingudes amb els resultats reals, on es mesurarà l'impacte dels diferents paràmetres per determinar el rendiment dels algoritmes usats, avaluar les diferències i treure'n conclusions.

Introducció

Un sistema recomanador es defineix com un mecanisme que, mitjançant la informació dels usuaris, ofereix productes del seu interès de forma automàtica.

La magnitud d'aquests sistemes és tan gran que els trobem en el nostre dia a dia pràcticament a qualsevol consulta que realitzem per Internet. Comprar un producte en una plataforma com Amazon, realitzar una cerca sobre algun tema concret a Google o decidir veure una pel·lícula o una altra a Netflix, nodreixen d'informació a aquests sistemes amb la finalitat de proporcionar a l'usuari una navegació més fluida i satisfactòria.

Seria impensable que un usuari a l'hora de comprar un producte, entre els milions de productes que s'ofereixen en el comerç electrònic, fos capaç de trobar el que més s'adeqüi a

les seves necessitats, o bé de trobar un vídeo concret a Youtube entre milers de similars.

Per citar un exemple, en el cas de Netflix, la companyia assegurava l'any 2012 que el 75% dels continguts que consumeixen els usuaris de la seva plataforma provenen d'algun tipus de recomanació. L'èxit de les recomanacions, segons la companyia, és gràcies a l'optimització contínua de l'experiència d'usuari, que es pot mesurar a partir del seu grau de satisfacció.

En general, no només els usuaris es beneficien d'aquests sistemes, també atorguen a les empreses la possibilitat d'oferir productes més variats i entendre millor què volen els usuaris i, en definitiva, vendre més.

Existeixen multitud d'algorismes d'aprenentatge automàtic. En aquest treball, realitzarem un estudi de les diferents tècniques que s'utilitzen, les analitzarem i valorarem la seva eficiència per extreure conclusions de cara a determinar si generen, o no, bones recomanacions.

1. Context

Abans d'entrar més en profunditat en el funcionament dels sistemes recomanadors per a pel·lícules, realitzarem una petita introducció al món dels sistemes recomanadors a nivell del seu funcionament general. Aquests necessiten:

- Un **conjunt d'usuaris** consumidors dels productes.
- Un **conjunt de dades** que descriu el contingut dels productes.
- Un **conjunt de puntuacions** que s'aconseguiran d'una manera o una altra depenent de la tècnica de recomanació que s'utilitzi.

Respecte al **conjunt d'usuaris**, la informació pot ser obtinguda per diverses vies diferents i depenen molt de la plataforma i dels productes que es vulguin recomanar.

En el nostre cas, si ens fixem en el sector de les pel·lícules, podem imaginar-nos que en les plataformes de vídeo més exitoses d'avui dia que tots coneixem, la informació s'obté de les nostres interaccions amb el servei, és a dir, si hem vist o no una pel·lícula i si hem pitjat el botó de "m'agrada" o el de "no m'agrada".

Un altre aspecte, a més de la informació dels usuaris, és la informació del **conjunt dels productes** dels quals disposem, en el nostre cas, les pel·lícules.

Sembla una afirmació massa òbvia però, per recomanar qualsevol cosa, hem de conèixer bé què és aquesta cosa. A tota la informació obtinguda de les nostres pel·lícules la denominarem “contingut”, nom basat en els sistemes recomanadors anomenats **sistemes basats en el contingut** (*content-based*) que basen la seva informació únicament en els aspectes dels mateixos productes que es recomanen.

Les característiques més típiques, en aquest domini, són:

- El/la director/a de la pel·lícula
- El gènere de la pel·lícula
- El títol de la pel·lícula
- Tot el repartiment d'actors
- L'argument de la trama

Aquests són simplement alguns dels paràmetres que es solen tenir en compte i els que més destaquen quan pensem en el contingut d'una pel·lícula però, de fet, podem tenir tants paràmetres com vulguem, per exemple:

- És una pel·lícula familiar?
- Apareix en Leonardo DiCaprio de jove?
- És una pel·lícula on apareixen el mateix nombre d'homes que de dones?

Com podem veure, el que podem considerar com a “contingut” pot ser literalment infinit.

Com hem mencionat prèviament, un últim aspecte que necessiten tots els sistemes recomanadors, sigui del tipus que sigui, és un **conjunt de puntuacions**.

Les puntuacions són, en última instància, **valors numèrics que s'associen a cada producte que vulguem recomanar**, de manera que, per exemple, si una pel·lícula obté una puntuació de 250 i una altra una puntuació de 110, sabrem ordenar-les perfectament de més a menys encertada per a un usuari en concret i podrem mostrar-les en l'ordre que ens convingui perquè l'experiència d'usuari en general sigui millor. Les puntuacions són importants de cara a valorar quin és el “graú d'aparellament” entre un usuari i cada una de les pel·lícules.

Els sistemes recomanadors es poden classificar en diferents categories segons la tècnica utilitzada per fer les recomanacions. Aquestes tècniques poden estar centrades en els usuaris, els productes, les puntuacions, o bé en una combinació d'aquests. Un cop mencionades, a l'apartat 2, Estat de l'art, presentarem les tres estratègies de recomanació més conegudes:

- **Popularity filtering**
- **Content-based filtering**
- **Collaborative filtering**

2. Estat de l'art

Inicialment, els sistemes recomanadors repliquen un funcionament que ja existeix i s'utilitza als sectors més tradicionals del màrqueting. Un exemple clàssic podria ser el fet d'anunciar productes escolars en un lloc proper a una escola, o al mes de setembre, data d'inici de l'any escolar i quan sorgeixen les necessitats de material precisament escolar.

Tenint en compte això, els primers sistemes recomanadors doncs, no aporten res nou ni innoven amb metodologies totalment noves, sinó que únicament es dediquen a imitar les tècniques de màrqueting ja existents a altres sectors i adaptar-los al comerç electrònic i als serveis en línia.

Un sistema recomanador bàsic, dels primers que van sorgir, és el basat en la popularitat. La lògica darrere aquests sistemes és senzilla perquè els productes que més es consumeixen solen ser els que la gent prefereix i, per tant, els millors a recomanar i anunciar.

Amb el pas del temps, augmenta massivament la quantitat de dades que podem adquirir dels usuaris que naveguen per Internet: es normalitza el funcionament de creacions d'usuaris, l'obligació de fer *login* per visualitzar continguts de pàgines web, les eines per puntuar i indicar si un contingut és del teu gust o no, etc. En aquest context, els sistemes recomanadors estan dotats de moltes més dades i es comencen a estudiar altres estratègies de recomanació més complexes com les estratègies basades en el contingut i les basades en el filtratge col·laboratiu.

Avui dia, els sistemes recomanadors segueixen en expansió continua i es comencen a usar sistemes basats en la Intel·ligència Artificial (IA) aplicant tècniques d'aprenentatge automàtic (*machine learning*) més avançades com l'aprenentatge profund (*deep learning*). L'aprenentatge profund està resultant molt útil per fer recomanacions en serveis com Youtube perquè li permet treballar a gran escala, sobre un **conjunt de dades dinàmic** i tractar una gran quantitat de factors externs que no es poden observar a simple vista. El sistema que usa Youtube consisteix en dues xarxes neuronals: una per generar candidats i una altra per rànquing [18].

Per exemple, amb la situació social actual respecte al virus de la Covid-19, el comportament i les conductes que tothom podia tenir com assentades i consolidades han estat modificades per complet. Un sistema recomanador basat en *collaborative-filtering*, fixant-se en la informació que prèviament els usuaris han generat, no podrà recomanar de manera tan òptima com un sistema basat en *deep learning* que, a l'estar contínuament adaptant-se, sabrà identificar que les conductes de les persones han canviat i, per tant, recomanar les noves necessitats sorgides d'un canvi social com aquest.

Aquestes tècniques, tot i que encara no son usades al 100% pel sector digital, són sens dubte les més prometedores i les que marcaran el futur del món dels sistemes recomanadors.

A continuació mostrarem una explicació més detallada de cadascuna de les estratègies més utilitzades en el sector digital.

2.1. Popularity filtering

Les estratègies basades en la popularitat (*Popularity filtering*) són les estratègies menys complexes i, alhora, les més fàcils d'implementar. Això no vol dir que no siguin efectives, tot al contrari, de vegades la millor recomanació que podem fer-li a un usuari és precisament el producte més venut a nivell global que, en el nostre cas, seria la pel·lícula amb més bona valoració dels usuaris (*user rating*) del nostre catàleg, la qual és molt possible que sigui del grat de la gran majoria d'usuaris.

2.2. Content-based filtering

Pel que fa a les estratègies basades en el contingut (*Content-based filtering*), es basen en aconseguir la informació del contingut que l'usuari mateix consumeix, en el nostre cas, la informació pròpia de cada pel·lícula. En aquest àmbit, la dades més rellevants solen ser: el gènere de la pel·lícula, el/la director/a, el repartiment d'actors/actrius principals i secundaris, i l'argument de la trama.

Aquest últim, l'argument, es representa mitjançant un conjunt de paraules clau o *keywords* que defineixen a grans trets les parts importants de la trama cinematogràfica; per exemple, per la pel·lícula de *Toy Story*, unes bones *keywords* podrien ser *joguina*, *cowboy* o *familiar*. És important remarcar que la paraula "animació", tot i que defineix la pel·lícula bastant bé, no seria una *keyword* ja que en aquest cas correspondria al gènere.

2.3. Collaborative filtering

Les estratègies basades en la "col·laboració entre usuaris" (*Collaborative filtering*), es basen en extreure la informació necessària per fer una recomanació a partir dels propis usuaris, en el nostre cas, dels consumidors de les pel·lícules.

Aquesta és sens dubte l'estratègia més complexa de totes les comentades anteriorment, no tan sols per la dificultat d'implementació, sinó també per la necessitat de tenir un conjunt de dades suficients (usuaris, pel·lícules i valoracions) perquè les recomanacions siguin efectives i valorables.

Tot i basar-se en l'extracció d'informació dels mateixos usuaris per igual, podem distingir entre dues fonts d'informació diferenciades. Segons la font d'informació, podem basar les nostres recomanacions amb una estratègia o

una altra d'acord al que interressi al nostre recomanador en particular. Les dues estratègies basades en *collaborative filtering* són: estratègia **item-based** i estratègia **user-based**.

2.3.1 Item-based Collaborative filtering

Aquesta estratègia de la tècnica *collaborative filtering* consisteix en predir quin ítem "s'ajusta" més per ser recomanat tenint en compte la similitud entre els ítems que han estat puntuats pels usuaris.

A diferència del mètode *content-based*, aquest no es fixa en la pròpia informació del ítem, és a dir, no comprovem el gènere o el director de la pel·lícula per exemple, si no que ens fixem, entre d'altres aspectes, en si als usuaris als qui els ha agradat una pel·lícula determinada X també els agrada una altra pel·lícula Y, i aquesta coincidència es produeix amb molta freqüència (a molts usuaris que els ha agradat X, també els agrada Y). Si això succeeix, podem determinar que aquestes dues pel·lícules tenen una similitud elevada ja que els usuaris així ho han decidit amb les seves valoracions.

2.3.2 User-based Collaborative filtering

Aquesta estratègia, dins la tècnica de recomanació de *collaborative filtering*, consisteix en predir quin ítem "s'ajusta" més per ser recomanat tenint en compte la similitud entre els usuaris entre si, és a dir, entre un usuari X i tota la resta d'usuaris que ens interressi valorar o que siguin similars a X.

Per fer ús d'aquesta estratègia, és molt important conèixer la informació dels usuaris mateixos, és a dir, els seus gustos. Existeixen diversos mètodes per classificar un usuari en un tipus o un altre, segons la seva edat, el seu sexe, la seva situació laboral, etc. Aquests, però, són mètodes que requereixen de la confirmació de l'usuari perquè les seves dades puguin ser utilitzades i, en molts casos, atempten directament la privacitat dels usuaris mateixos.

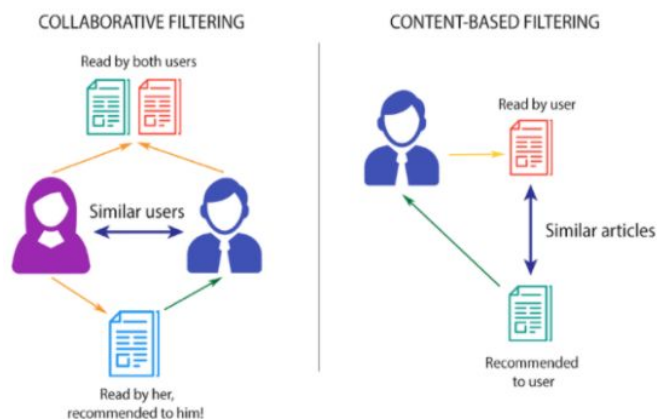


Fig. 1. Il·lustració que mostra les dues estratègies: collaborative filtering i content-based filtering

3. Formulació del problema

L'objectiu d'aquest projecte és el disseny i la implementació d'un sistema recomanador a partir de l'estudi i l'anàlisi de les diferents tècniques de recomanació existents. Com a consideracions prèvies, realitzarem un estudi de l'estat de l'art de les tècniques de recomanació més esteses i ens centrarem exclusivament en un domini: les pel·lícules.

L'objectiu de dissenyar i implementar un sistema recomanador de pel·lícules és el poder entendre la pipeline de processos que intervenen en la creació d'un sistema com aquest. Des de com s'obtenen i es tracten les dades de les quals disposem inicialment (obtenció, estructura, format, etc.) fins arribar a obtenir una recomanació.

Dins del possible, com a aportació pròpia al món dels sistemes recomanadors, aplicarem canvis subtils a les fórmules més usades per demostrar si hi ha alguna casuística en concret on, el canvi proposat, funciona millor que el sistema original i tradicionalment usat.

Finalment, per poder veure i utilitzar el sistema d'una manera més pràctica i visual, dissenyarem una petita interfície gràfica on puguem aplicar les mateixes comandes escrites pel terminal però d'una manera més interactiva. D'aquesta manera aconseguirem que per a un usuari amb un perfil menys tècnic, pugui entendre a nivell usuari els càlculs que s'estan realitzant i el perquè darrere de cada recomanació.

3.1 Solució del problema

S'han proposat 3 sistemes diferenciats, un per cada tipus d'estratègia principal: *popularity based*, *content-based* i *collaborative filtering*.

Segons l'estratègia a utilitzar, la informació que cal tractar del conjunt de dades d'entrada és diferent. També varien les mètriques necessàries per calcular les similituds entre ítems o entre usuaris. Es per això que ha calgut fer un estudi previ per entendre el comportament de cada mètrica i ser capaç de veure com introduir-hi modificacions per millorar l'eficiència del sistema.

L'objectiu final de totes les estratègies és comú: un llistat de quines són les pel·lícules que es recomanen, mostrades en ordre de més a menys pes (de més a menys recomanables). A partir dels resultats de cada estratègia, que segurament no coincidiran, s'ha donat una explicació del perquè hi ha diferències i si es pot dir que una estratègia es "millor" que una altra.

4. Gestió del projecte

Un dels principals problemes que han sorgit a l'hora de començar a donar contingut al projecte ha estat el de l'obtenció de les dades.

En un primer moment, l'objectiu era crear una pàgina web en la qual els diferents usuaris a qui els donéssim accés, poguessin introduir els seus gustos, marcar d'alguna manera el botó de "m'agrada" a unes quantes pel·lícules i, d'aquesta manera, anar enriquint la nostre pròpia base de dades. En definitiva, aconseguir un conjunt d'usuaris, un conjunt d'ítems (pel·lícules) i les seves valoracions. Aquest objectiu va ser descartat finalment ja que els possibles inconvenients eren massa grans (temps disponible molt limitat, d'octubre de 2020 a gener de 2021).

Bàsicament, és necessari comptar amb un elevat nombre d'usuaris, disposats a omplir certes dades (de l'ordre de milers), i informació sobre pel·lícules (de l'ordre de milers) perquè el conjunt de dades tingui prou entitat com per poder obtenir recomanacions.

La solució emprada ha estat utilitzar dades ja creades. La gran majoria de dades han estat extretes d'IMDB (Internet Movie DataBase), ja que disposa d'una gran varietat d'informació, així com conjunts de dades de diferents tamanys que ens aniran molt bé de cara a realitzar les proves pertinents previstes per aquest projecte.

Respecte la metodologia

Les diferents fases del projecte es mostren a la Taula 1.

Fase	Duració	Pes (%)
Fase Inicial (Fase 1)	17 dies (19 h)	5%
Fase d'Anàlisi (Fase 2)	41 dies (114 h)	30%
Fase de Desenvolupament (Fase 3)	103 dies (133 h)	35%
Fase Final (Fase 4)	13 dies (38 h)	10%
Fase de documentació i reunions de control (Fase General)	120 dies (76 h)	20%
TOTAL:	120 dies (380 h)	100%

Taula 1. Taula del sumatori i distribució de les hores dividit en fases

S'ha implementat un diagrama de Gantt on s'ha volgut mostrar una representació gràfica de les fases on podem observar més fàcilment el *workflow* general entre fases. [Consultar la documentació per la visualització del diagrama.]

Respecte la gestió econòmica

Per la identificació de costos i l'estimació d'aquests, el càlcul aplicat ha estat un simple sumatori de tots els costos de recursos humans, materials (hardware) i de programari (software) utilitzats.

Donada la situació social actual respecte al virus de la Covid-19, la manera de realitzar reunions i consultes de

forma remota és obligada i, per tant, disposar d'eines per poder comunicar-se virtualment és essencial, les quals han estat sumades al cost de material. Aquests càlculs queden reflectits en la Taula 2.

Recursos	Preu total
Humans	5.783,61€
Materials	1.999,98€
De programari	7,96€
TOTAL:	7.791,55€

Taula 2. Taula dels costos totals

5. Disseny i implementació

S'han proposat 3 sistemes diferenciats, un per cada tipus d'estratègia principal: *popularity based*, *content-based* i *collaborative filtering*.

5.1. Software utilitzat

Per a la realització de la implementació de cadascuna de les estratègies, amb l'objectiu de mesurar l'eficiència de cada un dels algorismes i veure com es comporten, utilitzarem el llenguatge de programació Python (versió 3.8.5) amb el complement d'algunes llibreries utilitzades comunament en problemes d'aprenentatge automàtic com *pandas*, *sklearn* i *numpy*.

De cara a la implementació de la interfície gràfica, l'eina utilitzada ha estat Flask.

5.2. Dades

La majoria de les dades utilitzades al llarg de la implementació han estat extretes de <https://grouplens.org/>. Més concretament, del projecte que duen a terme anomenat *MovieLens*.

Els arxius o *datasets* contenen extraccions de tot tipus d'informacions recollides d'usuaris i pel·lícules al llarg del temps. Les diferenciarem segons el tamany del *dataset*.

Per alguna prova en concret, s'han utilitzat diferents arxius, paràmetres o dades, però en la gran majoria podem dir que la informació usada està formada bàsicament pels següents arxius:

- **movies_metadata.csv** → La informació més rellevant que conté és la del títol original de la pel·lícula, el resum de la trama (sinopsi) i la quantitat i la mitjana de vots comptabilitzats
- **keywords.csv** → Conté una paraula o una successió de paraules clau relacionades amb una única pel·lícula.
- **credits.csv** → Conté informació del repartiment d'actors.
- **u.data** → Conté la informació de les valoracions de les pel·lícules per cadascun dels usuaris.

5.3. Recomanació basada en la popularitat

Habitualment trobarem aquest tipus de recomanacions en les plataformes comercials sota el títol de “Les més vistes”, “Les que més agraden a tothom” o “Les que més nota han rebut”.

Cal establir un nombre de vots mínims perquè no distorsionar el rànquing final i sigui el més fidel possible al que realment volem obtenir, que és el *top-20* de les pel·lícules amb millor puntuació del nostre catàleg de pel·lícules.

Els resultats obtinguts estan basats en el càlcul del *weighted rating*, que és el mètode que s'utilitza per a qualsevol recomanació basada en la popularitat. El *weighted rating* aplica la següent expressió matemàtica:

$$\text{WeightedRating}(\text{WR}) = \left(\frac{v}{v+m} \cdot R \right) + \left(\frac{m}{v+m} \cdot C \right)$$

On:

- **v** és el nombre de vots per la pel·lícula
- **m** és el mínim de vots requerits per a ser llistada
- **R** és la puntuació mitjana de la pel·lícula
- **C** és la puntuació del vot mig d'entre totes les dades de què disposem

5.4. Recomanació content-based

Habitualment, trobarem aquestes recomanacions en les plataformes comercials sota el títol “Perquè t'han agradat pel·lícules semblants” o bé “Si t'ha agradat una pel·lícula en concret també t'agradaran aquestes”.

El sistema està implementat perquè ens mostri 2 outputs diferents. Per una part, inicialment, ens mostra la llista només tenint en compte l'arxiu *movies_metadata.csv*. Recordem que aquest arxiu conté dades com el gènere i el plot de la pel·lícula, entre d'altres de menys rellevància.

Per altra banda, la segona part de l'execució del sistema recomanador *content-based*, ens mostra un llistat de també 10 pel·lícules, però en aquest cas, se li sumen a l'equació els fitxers *keywords.csv*, que conté les paraules clau, i *credits.csv* que conté informació del repartiment d'actors. Hi ha diversos mètodes per calcular les recomanacions basades en el contingut. En aquest projecte s'ha usat la fórmula següent:

$$\cos(\mathbf{x}, \mathbf{y}) = \frac{\mathbf{x} \cdot \mathbf{y}^T}{\|\mathbf{x}\| \cdot \|\mathbf{y}\|} = \frac{\sum_{i=1}^n x_i \cdot y_i^T}{\sqrt{\sum_{i=1}^n (x_i)^2} \sqrt{\sum_{i=1}^n (y_i)^2}}$$

Aquesta fórmula és coneguda com la **similitud cosinus** i s'usa per calcular numèricament la similitud entre dos documents i que, aplicada al nostre context, correspon a la similitud entre dues pel·lícules.

5.5. Recomanació collaborative filtering

Habitualment, trobarem aquestes recomanacions en les plataformes comercials sota el títol “Perquè ha agradat a usuaris semblants a tu” o “Usuaris similars també miren:”.

Per a l'estratègia *user-based*, la matriu de similitud d'usuari es calcula a partir d'una mètrica que mesura la similitud entre dos usuaris. D'igual manera, per a l'estratègia *item-based*, la matriu conté la similitud entre un parell de pel·lícules.

Una mètrica comunament utilitzada és la similitud cosinus, mètrica que ja hem utilitzat prèviament per al càlcul de l'estratègia *content based*. La similitud cosinus ens retorna un valor entre 0 i 1, en el nostre cas, ja que no disposem de dades amb *ratings* negatius.

Un cop tenim la nostra matriu de similitud, podem començar a predir els *ratings* que no han estat inclosos en les dades, és a dir, començar a recomanar pel·lícules a usuaris que no coneixen aquestes pel·lícules en qüestió.

Per a *user-based filtering*, la predicció que utilitzem per determinar el *rating* de l'usuari u cap a una pel·lícula p ve donada per la suma de tots els vots d'altres usuaris cap a la pel·lícula p , on el pes de cada vot dels altres usuaris es determina per la similitud cosinus entre cada usuari amb l'usuari u . Classifiquem els *top-k* usuaris o pel·lícules més semblants, sent k un valor parametrizable.

Una última estratègia emprada per millorar la qualitat de les recomanacions, a part de predir els *top-k*, ha estat el tractar de manera diferent els vots provinents dels usuaris “extrem”. Definim com a un usuari “extrem” aquell usuari que sempre tendeix a donar vots massa baixos o massa alts a totes les pel·lícules. Per tal de relativitzar el pes d'un usuari extrem, podem suposar que el més convenient no és la puntuació absoluta que finalment ha donat a una pel·lícula sinó “ponderar” les seves valoracions en funció de la seva mitjana de vots. Per aconseguir això, primer cal calcular la mitjana de vots de cada usuari i després recalculer cada una de les seves valoracions restant-li la mitjana dels vots que ha donat.

5.6 Interfície gràfica

Les interfícies gràfiques són una simple traducció dels codis Python amb les mateixes funcionalitats que hem vist en els apartats que descriuen les estratègies implementades. La diferència principal entre les dues és simplement estètica. En lloc de veure els resultats via terminal, els podrem veure via web (HTML)

6. Model d'avaluació

Un cop vistes les implementacions, necessitem mesurar d'alguna manera com sabem si estem realitzant bones recomanacions i, a més, saber què podem considerar com a una bona recomanació en un primer moment.

Per determinar si una recomanació és bona o no l'hauríem de comparar amb les puntuacions que realment donen els usuaris. D'alguna manera seria mesurar l'error de la predicció feta pel sistema recomanador. Amb aquest objectiu, descriurem algunes mètriques basades en la mesura de l'error, però també s'inclouran altres mètriques per avaluar la qualitat de les recomanacions d'una forma alternativa.

A continuació, fem una breu descripció de les mètriques, anomenades mètriques d'exactitud predictiva, que mesuren com de prop estan les puntuacions predites pel sistema recomanador respecte a les puntuacions reals fetes per l'usuari.

- **Mean Absolute Error (MAE)** → És la mitjana de la diferència entre el valor predit pel recomanador i el valor donat per l'usuari (*rating*). Només volem saber la diferència entre el valor predit i el real.
- **Mean Squared Error (MSE)** → L'error quadràtic mig és similar al MAE però, en aquest cas, elevem a 2 (agafem el quadrat) de l'error absolut. D'aquesta manera aconseguim que les diferències (errors) més grans tinguin encara més pes. És una manera de penalitzar justament les diferències (errors) més grans perquè s'elevem al quadrat.

A diferència de les mètriques d'exactitud predictiva (MAE i MSE), les mètriques següents ens ajuden a determinar en quina mesura el recomanador ha estat útil a l'hora d'assistir a l'usuari i ha encertat recomanant certes pel·lícules i evitant-ne d'altres.

- **Precision** → És el nombre d'ítems **rellevants** que el sistema ha recomanat respecte el nombre total d'ítems que el sistema ha recomanat.
- **Recall** → És el nombre d'ítems rellevants que el sistema ha recomanat respecte el nombre total d'ítems rellevants.

Els mètodes que hem vist ens ajuden a entendre l'eficiència general del resultat que obtenim a partir del sistema recomanador, però no ens aporta informació de com els ítems han estat ordenats. Un model pot tenir un bon valor de MSE o MAE però si les primeres recomanacions no són rellevants per l'usuari, no ens és útil. Els mètodes següents ens ajuden a solucionar aquest problema.

- **Discounted Cumulative Gain (DCG)** → Classifica entre ítems rellevants, és a dir, distingeix entre un ítem rellevant i un altre, i situa a un per sobre de l'altre, tot i que els dos no deixen de ser rellevants. Inicialment, penalitza els ítems amb menor rellevància dins dels rellevants i, finalment, dóna un valor que ens indica l'èxit de la recomanació.
- **Mean Reciprocal Rank (MRR)** → Es centra en localitzar el primer ítem més rellevant. El valor MRR d'un sistema amb el primer element més rellevant a la posició 3, serà més alt que el d'un sistema amb l'element més rellevant a la posició 4.

- Novetat → En algunes plataformes, trobem en molts casos l'apartat de “novetats”. En dominis com, per exemple, la música, és una bona mètrica a tenir en compte.
- Diversitat → La diversitat és un altre aspecte a valorar, si hi ha una alta varietat en el nostre model, voldrà dir que sempre tindrem pel·lícules diferents per veure.
- Mètriques empresarials → Sens dubte, de les més importants al món on vivim. En última instància, una companyia sempre busca un model que generi beneficis o ajudi a complir els seus objectius empresarials.

7. Conclusions

Després de tot l'estudi, la implementació i la validació de les dades, i l'anàlisi dels resultats podem dir que tots 3 algoritmes base (*popularity filtering*, *content-based filtering* i *collaborative filtering*) tenen la capacitat de proporcionar bones recomanacions.

El terme “bona recomanació”, tot i disposar d'eines de mesura que ens indiquen els càlculs de l'error en les prediccions, dependrà molt de l'usuari que estigui sent recomanat.

Per exemple, un usuari al qual se li recomanem una sèrie de pel·lícules que s'adeqüen bastant als seus gustos però, les dues primeres pel·lícules no són exactament les que vol veure, molt possiblement acabarà pensant que la recomanació que li ha donat el sistema no es bona.

En molts casos. la tècnica a escollir per un sistema recomanador dependrà de les preguntes que ens puguem fer i sobre el que vulguem resoldre.

També podem determinar que, en estar tractant amb algoritmes que aprenen directament de les dades amb les quals disposem, la quantitat i qualitat de les dades influeixen directament en els resultats que s'obtenen. Donat això, els sistemes que obtenen les dades mitjançant estratègies de *deep learning* es presenten molt prometedores de cara al futur.

Si disposem d'un volum de dades suficient sobre els diferents usuaris de la nostra plataforma, una estratègia *collaborative filtering* serà sempre una garantia que el que estem recomanant són les pel·lícules que els usuaris (que se'ns assemblen) estan consumint més i, per tant, tindrem poc marge a equivocar-nos en la nostra recomanació.

Si pel contrari no disposem d'un gran volum de dades dels usuaris per alimentar el nostre sistema recomanador, no podem garantir que els resultats obtinguts siguin els esperats o desitjats amb *collaborative filtering*, per tant, haurem de buscar altres estratègies com la *content-based filtering*, on el filtratge dependrà del contingut mateix de les pel·lícules.

Tot i conèixer quin tipus d'estratègia funciona millor per cada cas, sempre serà una bona pràctica el combinar les nostres estratègies de manera que l'usuari pugui tenir suficients eleccions rellevants de pel·lícules a veure. Per exemple, una estratègia de popularitat simple, basada únicament en la puntuació que té una pel·lícula a IMDB o basada en la quantitat de “m'agrada” que hagi obtingut la pel·lícula en qüestió, proporcionarà al nostre sistema un salt de qualitat en quant a les recomanacions realitzades sense haver necessitat grans quantitats d'informació.

Com hem comentat amb anterioritat, per a plataformes com Netflix que disposen de molts milions d'usuaris i, una part molt elevada dels continguts que es visualitzen venen donats per les recomanacions, és imprescindible que aquests sistemes i estratègies estiguin altament depurats i siguin molt eficients, ja que hi ha una clara i directa relació entre les recomanacions i els ingressos que genera la companyia.

A nivell personal, aquest treball m'ha ajudat a entendre millor quines són les eines que les grans empreses utilitzen per a recomanar productes, pel·lícules o publicitat variada per procurar que consumeixis el producte en qüestió.

Puc dir que aquest treball m'ha canviat la manera de fixar-me en la publicitat que rebo i en les recomanacions que em fan les plataformes i navegadors que uso habitualment. Suposo que com jo mateix, tots hi estem extremadament exposats diàriament en qualsevol moment que passem a Internet.

Agraïments

Vull agrair especialment a la meva tutora Neus Català Roig per tot l'esforç, l'ajuda i el constant seguiment del projecte al llarg d'aquests mesos. Estic convençut que els seus consells i treball han estat la clau per a la realització del treball en la seva totalitat.

Referències

- [1] Com funcionen els sistemes de recomanació (Netflix/Amazon):
https://www.youtube.com/watch?v=n3RKsY2H-NE&ab_channel=ArtoftheProblem
- [2] Divulgació respecte als sistemes recomanadors:
https://www.youtube.com/watch?v=Eeg1DEeWUJA&ab_channel=CS50
- [3] Com recomana Netflix les pel·lícules? Factorització de Matrius:
https://www.youtube.com/watch?v=ZspR5PZemcs&t=1s&ab_channel=LuisSerrano
- [4] Llista dels mètodes per a avaluar un sistema recomanador:
<https://towardsdatascience.com/an-exhaustive-list-of-methods-to-evaluate-recommender-systems-a70c05e121de>
- [5] Evaluant els sistemes recomanadors | Stanford University:
https://www.youtube.com/watch?v=VZKMyTaLI00&ab_channel=ArtificialIntelligence-AllInOne
- [6] Divulgació de l'avaluació dels sistemes recomanadors
https://www.youtube.com/watch?v=qG0wUgsEugw&ab_channel=WayfairDataScience

- [7] Avaluant mètriques per a sistemes recomanadors:
<https://towardsdatascience.com/evaluation-metrics-for-recommender-systems-df56c6611093>
- [8] Com construir un sistema recomanador de pel·lícules content-based:
<https://towardsdatascience.com/how-to-build-a-content-based-movie-recommender-system-92352f5db7c6>
- [9] Pàgina principal d'IMDB. Llistat de les pel·lícules més votades:
<https://www.imdb.com/chart/top/>
- [10] Imatge Content-based vs Collaborative filtering:
https://www.researchgate.net/figure/Content-based-filtering-vs-Collaborative-filtering-Source_fig5_323726564
- [11] Pàgina web oficial de la llibreria Python Flask:
<https://flask.palletsprojects.com/en/1.1.x/>
- [12] Pàgina web oficial de GanttProject:
<https://www.ganttproject.biz/>
- [13] Pàgina web oficial de Python:
<https://www.python.org/>
- [14] Pàgina web oficial de Scikit-learn:
<https://scikit-learn.org/stable/>
- [15] Pàgina web oficial de la llibreria Python Pandas:
<https://pandas.pydata.org/>
- [16] Pàgina web oficial de la llibreria Python NumPy:
<https://numpy.org/>
- [17] Pàgina web oficial de la llibreria Python matplotlib:
<https://matplotlib.org/>
- [18] Deep Neural networks a les recomanacions de Youtube:
<https://static.googleusercontent.com/media/research.google.com/ru//pubs/archive/45530.pdf>