

Collaborative Filtering (CF)¹

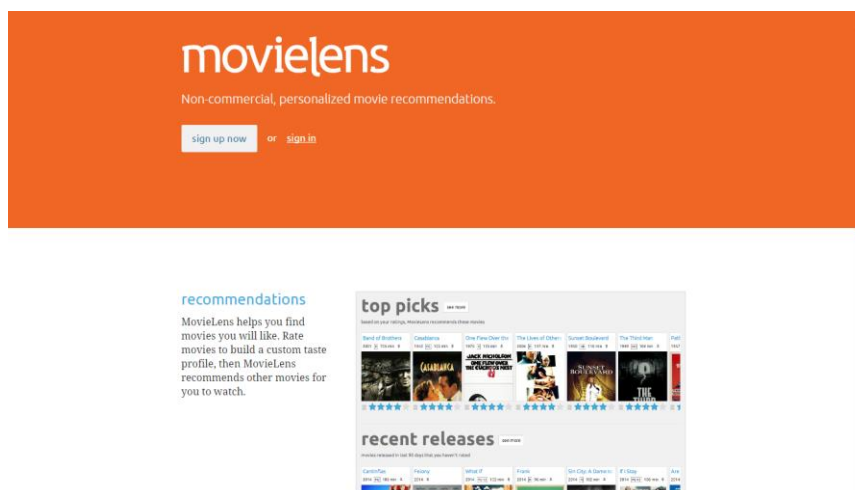
A köznyelv nemigen ismeri magát a szakkifejezést, amikor viszont „*közösségek által meghatározott véleményalkotás*”-ról vagy honlapokon „*személyes ajánlások*”-ról esik szó; a sok felhasználó már képes körülhatárolni a kifejezés jelentését.

A „**közösségi szűrés**” (**collaborative filtering - CF**) nem mondható újnak az emberiség történelmében. Lényegében „collaborative filtering” minden olyan eljárás, amely során egy-egy közösség tagjai közösen határozzák meg, hogy mi hasznos, mi haszontalan és mi kerülendő.

Legegyszerűbb változatát az emberség már a történelem előtti időkben is használta. Őseink bevált gyakorlata egy addig nem látott bogyó vagy gyökeres növény megtalálása esetén az volt, hogy a törzs egyes tagjai (feltehetően a gyengébb fizikai állapotban lévők ...) kóstolták meg azt először, s ha ők nem lettek betegek tőle, akkor azt a törzs többi tagja is megehette.

A XXI. században a különbség annyi, hogy a „**kóstoláshoz**” információs technológiát (IT) és *ajánlórendszereket* használunk.

Az első közösségi felhasználású, internetes technológián alapuló collaborative filtering eljárást használó weboldalak között talán a legismertebb a MovieLens (movielens.org), a személyes moziajánló, melynek példája jól szemlélteti a collaborative filtering működési elvét.



A movielens.org moziajánlója 2016-ban: www.movielens.org

¹ RIEDL–KONSTAN–MAJÓ–RÉVÉSZ: SZÁJTPROPAGANDA CÍMŰ KÖNYVÉNEK ÉS KERTMEGI: EGY AJÁNLORENDSZER SZEREPE A JÖVŐ MARKETINGJÉBEN OTDK DOLGOZAT ALAPJÁN KÉSZÜLT OKTATÁSI SEGÉDLET

A rendszer használatához egy regisztrációra van szükség, ahol a felhasználónak értékelni kell korábban látott filmeket egy skála segítségével. Ezek után **a rendszer az értékeléseket, egybeveti más felhasználókéval**, és megpróbálja felhasználókat összehasonlítani. Lényegében nem történik más „csak” a rendszer a mozifilmek tetszési indexe alapján megkeresi a **legközelebbi vélemény szomszédunkat**, akivel legjobban hasonlítunk egymásra (ugyanazokra a filmekre ugyanolyan értékelést adtunk).

A profil-összehasonlításhoz informatikai környezetben valamilyen egzakt skála és többváltozós statisztikai számítási program használata szükséges (és komoly számítási kapacitás, ha több ezer felhasználóról van szó) Egy ilyen rendszerben az adatok összevetése két dimenzió mentén történik. Az *átfedés* ebben az esetben azon filmek száma, amelyeket mindkét felhasználó értékelt, a *korreláció* pedig azt mutatja, hogy például az adott filmek esetében mennyire hasonló módon ítéltük meg azokat a filmeket, amelyeket mindketten láttunk.

Ideális esetben, ha az adott rendszernek kelő számú felhasználója van találhatók olyan felhasználói csoportok, akikkel összevetve az átfedés és a korreláció egyaránt magas. Ha a korreláció magas értéke (1) arra utal, hogy egy véleményen vagyunk. A magas átfedés pedig azt jelenti, hogy az egyetértésünk nem csupán véletlenszerű, hiszen nagy mennyiségű adaton alapul.

A *collaborative filtering* algoritmusok melyek az adott felhasználónak ajánlanak többféle informatikai megoldással párosulhatnak. Léteznek manuális collaborative filtering rendszerek, amelyben emberek tesznek vagy kérnek egyéni ajánlásokat, azonban **a legtöbb kereskedelmi alkalmazás automatizált rendszereket használ, amelyek vevőpreferenciákat gyűjtenek, tárolnak, elemeznek, és megkeresik a hasonló ízlésű vevőket**, és tapasztalataikat felhasználva ajánlanak másoknak bizonyos termékeket.

Az ajánlórendszerek két meghatározó irányzata fejlődött ki az évek során: a **termék és tartalom alapú**, ill. a **felhasználó és preferencia** alapú rendszerek. *Tartalom alapú ajánlórendszerek* esetén a szűrőfeltétel maga a termék és a szolgáltatás. Ilyenek például az online könyvtárházak (pl. Bookline.hu), ahol könyvcím, szerző vagy épp témakör szerint is kaphatunk ajánlásokat: ilyenkor például a már kosárba tett szerzőtől ajánl a rendszer másik könyvet is. Ezek a rendszerek korábbi vásárlásainkból indulnak ki, azok alapján feltételeznek jövőbeli pozitív

tetszésindexünket.

A *preferencia-alapú rendszerek* viszont a termékjellemzők helyett, az előzőleg azonos termékeket vásárló vevők hasonló preferenciáira hagyatkoznak. Így aszerint teszik meg ajánlásait a jövőben megvásárolandó termékre, hogy „*közeli szomszédaink*” (aki a miénkhez elég hasonló ízléssel rendelkezik) milyen terméke(ke)t vásárolt már meg, illetve egy adott termékről vagy szolgáltatásról milyen véleménnyel van.

A **collaborative filtering algoritmusok** nemcsak a felhasználó által megadott direkt információkat (pl.: értékelések), hanem indirekt adatokat is feldolgozhat. A korszerű ajánló rendszerek már a felhasználók külön interakcióját, beavatkozását sem igénylik, hiszen ajánlásait a cookie-k és aktív algoritmusok segítségével, teljesen *automatizáltan* juttatják el hozzánk. A számítógépünkre telepített cookie-k segítségével – mely követi a digitális lábnyomunkat - a tárolt információk hasonlósága alapján találják meg a hozzánk minél inkább hasonló preferenciákkal (ízléssel) rendelkező „*legközelebbi szomszédaink*”-at, és ad ajánlásokat.

Az **automatizált collaborative filtering** három különböző fajtáját tudjuk megkülönböztetni, amik három egymástól eltérő, de kompatibilis rendszerhez vezetnek. Ezek a *pull-active CF*, a *push-active CF*, valamint az *automatizált CF*. A három rendszer különböző feladatokat hajt végre, ezért ugyanazon a weboldalon akár egyszerre mindhárommal is találkozhatunk.

A *pull-active CF* esetén a felhasználó – a névből is adódóan – aktívan részt vesz az ajánlattétel folyamatában. Ezt azáltal teszi, hogy a rendszertől „lekérdezéseket” kezdeményez (pl. „*Szeretném megtudni, milyen márkájú cipőt vett Gábor, mert amit ő vesz, az biztos csak jó lehet.*”).

A *push-active CF*-et a Lotus Research alkalmazottai, David Maltz és Kate Ehrlich fejlesztették ki, nem sokkal elődje után. Ez a rendszer a későbbi nagyszabású információvezérek megszületésének kedvezett, mivel számukra azt tette lehetővé, hogy egy érdekes képet, hírt immár bárkinek „átolhassanak”, akiről úgy gondolták, hogy őt is érdekelheti. Az említett *aktív CF rendszerek* beavatkozási igényük révén, értelemszerűen, olyan közösségekben alkalmazhatók leginkább, amelynek tagjai már ismerik egymást, tisztában vannak egymás ízlésével.

Az automatizált CF rendszerek a korábbi interakciók ((pl. személyes értékelések, feed-back, stb.)) és a **cookie-k által közvetített adatokat felhasználva állapítják meg a személyre szabott ajánlást**. Utóbbi rendszerek végig követnek bennünket az egyes weblapokon tett

értékeléseink során, megjegyzik azoknak eredményeit, majd a fogyasztói ízlések hasonlóságai alapján kiszámítják, hogy mely más fogyasztók képesek a mi ízlésünket megjósolni. Mivel azonban az algoritmus a fogyasztással kapcsolatos, személyes adataink (nem, életkor, lakhely, végzettség stb) ebben a rendszerben nem relevánsak, akár rejtve is maradhatnak. Másképp mondva, legközelebbi véleményyszomszédunk lehetséges egyáltalán nem hasonlít ránk ☺ nem ismeri a nyelvünk és sosem fogunk találkozni vele az utcán ☺

Ma már nem megy meglepetészámba, ha egy filmsorozat megszállott rajongójaként pl. az Amazon-on megjelenik előttünk eladásra kínálva az imádott főszereplőnk nevével fémjelzett termék, de azon sem, ha pl. a Vaterán könyvet vásárolva a termékajánlóban nemcsak újabb könyvek, de akár horgászboto, vagy vászonképek is helyet kapnak.

A mai ajánlatokban ugyanis már olyan termékek is megjelenhetnek, amik ugyan szorosan nem köthetők az adott termékhez vagy annak felhasználási/fogyasztási tevékenységhez, egyéni preferenciáinkhoz viszont annál inkább.

az Amazon paradox termékkapcsolása: a Jack Bauer-baba

Ellenőrző kérdések:

a, Mi a magyar megfelelője a collaborative filteringnek (CF)?

b, Kit illetve kiket keres a CF algoritmus? Mi az eljárás célja?

c, Milyen adatokat dolgoz fel egy automatizált CF rendszer?

Adaptáció:

Készíts el egy 1 oldalas anyagot olyan általad ismert weboldalról, ahol láthatóan CF eljárást alkalmaznak. Mutasd be és elemezd az ajánlásokat, írd le véleményed azzal kapcsolatban, szerinted hogyan lehetne még fejleszteni a rendszeren?