

Tartalomajánló rendszerek

Kis-Király Máté



Netflix Prize

- ▶ Netflix: online filmkölcsönző és stream szolgáltató
- ▶ A verseny:
 - ▶ Célja a Netflix ajánlási hatékonyságának javítása 10%-kal
 - ▶ Díj: 1 millió dollár
 - ▶ Feladat: felhasználók addigi értékelései alapján megjósolni, hogy mi tetszik nekik (azok közül, amit még nem láttak).

Netflix Prize és a BME

▶ BME-TMIT csapata

- ▶ Gravity
- ▶ A ranglistán első hely
- ▶ A tesztalmazon holtverseny
- ▶ 20 perc késés -> 2. hely

Miért van szükség ajánlásra?

▶ Tartalom áradat

- ▶ Túl sok elem, túl sok választási lehetőség
- ▶ YouTube videók, Amazon könyvek, Netflix filmek

▶ Fogasztó

- ▶ Nem elég képzett
 - ▶ Pl. nagymama számítógépet szeretne
- ▶ Nem tudja, hogy mit akar
 - ▶ Pl. első alkalom egy kínai étteremben
- ▶ A tartalom folyamatos publikálása követhetetlen
 - ▶ Pl. Facebook post, Twitter tweets

Miért van szükség ajánlásra?

► Szolgáltató

- Hogy kezelje
 - a tartalmat,
 - a kapcsolódó meta adatokat,
 - a tranzakciós adatokat?
- Mi fontos az adatok közül?
- Milyen információt rejt?
- Hogyan nyerje ki a fontos információkat?
- Hogyan tud üzleti előnyt elérni?
 - Olyanokat ajánlani, amit a felhasználó szeretne
 - Ennek mi az ára

6

Meta adat: bármilyen adat, ami a tartalomhoz kapcsolódik. Pl. filmek esetén a rendező, a főszereplő, a műfaj.

Tranzakciós adat: felhasználó és az elemek közötti interakció során keletkező adat. Pl. IPTV rendszerben on-demand filmek értékelése, kedvencek közé helyezése egy-egy tranzakciós adat.

Fontos az, hogy mikor értékelt a felhasználó? Fontos, hogy az esti szappanoperát megnézi?

Ha megnézett két horror filmet, vajon ajánlható-e neki még több?

A filmekhez online forrásokból további információt lehet beszerezni, de ez kifizetődő?

A filmhez tartozó felirat sok információt rejt magában, de érdemes átnézni?

Tartalomajánlás

- ▶ Felhasználó érdeklődési körébe tartozó elemek ajánlása
- ▶ Egy függvény
 - ▶ Bemenet: adott felhasználó, adott elem
 - ▶ Kimenet: pontszám
- ▶ $R : \text{Felhasználó} \times \text{Elem} \rightarrow \text{pontszám}$
- ▶ Az R függvény értékét meghatározza minden elemre
- ▶ A legnagyobb értékű R -hez tartozó elem a „legjobb” javaslat

Elem: a javaslat tárgyára utal. A felhasználási környezettől függően lehet pl. zene, videó, film, termék egy web áruházban.

A felhasználó-elem összerendelés alapján megkülönböztetünk különféle ajánlási módszereket.

Nem személyre szabott ajánlás



Tulajdonképpen az előző definíció. Maga az ajánlás a legegyszerűbb esetben független az ajánlott elemek típusától, – szakkifejezéssel élve a domaintól – lehet ez a termék épp film, zenész, ruha stb. (de a tartalom elemek ugyanabból a domainből kerülnek ki). Az ajánlórendszer különböző módszerek szerint pontszámokat rendel az adott felhasználóhoz rendelhető elemekhez, majd a pontszám alapján rangsorolást végez, amely sornak az elején lévő elem kerül ajánlásra. Ebben az esetben a tartalom elemekről különböző információk állnak a rendszer rendelkezésére, ami alapján a rangsorolást és a javaslatot megvalósítja. Ilyen leíró információ például (filmek esetében) a műfaj, kategória, kiadás éve.

A nem személyre szabott módszer úgy tesz ajánlatot a felhasználónak, hogy előre definiált, felhasználó-független szabályokat vesz figyelembe. Ez történhet például úgy, hogy az adott domainben lévő legújabb, vagy legfelkapottabb, legnépszerűbb elemet kínálja fel. Érezhető, hogy ez a megoldás nem túl hatékony, hiszen minden felhasználó ugyanazt az ajánlási listát fogja kapni, függetlenül az ő érdeklődési körétől.

Személyre szabott ajánlás



Személyre szabott megvalósítás esetén a rendszernek van információja a felhasználóról is, és ezt felhasználva tud határozottan az ő igényeihez igazodva javaslatot tenni.

Látható, hogy a felhasználók halmazán kívül felhasználói információ is kerül az ajánlórendszerbe. Ez az adat olyan leíró információkat tartalmaz a felhasználóról, mint például kor, nem, lokáció. A rendszer működése során a felhasználói leírók alapján hozott szabályok szerint személyre szabottan tesznek ajánlatot. Ilyen szabály lehet például, hogy középkorú férfiak számára érdemes lehet drága alkoholt ajánlani. A perszonalizációs módszer segítségével sokkal pontosabb, a felhasználó érdeklődési körébe jobban illő ajánlások készíthetők.

Interakció alapú személyre szabott ajánlás



A paradigma kiinduló ötlete az, hogy a felhasználó kapcsolatot létesít a különböző tartalom elemekkel, amelyből olyan információk nyerhetők ki, amelyek tovább növelhetik az ajánlás hatékonyságát. Az interakció során a felhasználó értékelheti a terméket, vagy akár egy hosszabb ismertetőt is közölhet. Ezzel az újonnan szerzett információval kiegészült rendszer egyszerűsített ábrája látható a dián.

A kinyert információk nem alkotnak homogén adathalmazt, nem képviselnek azonos értéket. Megkülönböztetünk explicit és implicit adatot (a teljesség igénye nélkül): explicit, amikor a felhasználó közli a preferenciáit:

- a felhasználó értékeli az elemet (pl. 1-6 közötti érték)

- a felhasználó rangsorolja az elemeket

- két elem közül kiválasztja a számára megfelelőbbet

implicit, amikor kikövetkeztethető a felhasználó preferenciája, de ő maga ezt nem támasztja alá:

- milyen elemeket nézett meg

- hányszor nézte meg az elemeket

- miket vásárolt, mikre fizetett elő

- miket tett a kedvencek közé

Általánosságban elmondható, hogy az explicit adatok értékesebbek, mint az implicit, azonban sokkal nehezebben hozzáférhető és lényegesen kisebb mennyiségben áll a rendszerek rendelkezésére. A paradigma az ajánlást a felhasználók viselkedéséből, az adott visszajelzések alapján képes elkészíteni sokkal személyre szabottan, mint az eddigi megközelítések.

Kontextus alapú személyre szabott ajánlás



Az eddigi megoldások során két fajta entitást vettek figyelembe a rendszerek: a felhasználót és az ajánlható elemeket. A környezet alapú paradigmának a lényege, hogy figyelembe veszi az ajánlás idejében a felhasználó szűkebb, illetve tágabb környezetét (kontextusát), abból nyer ki bizonyos információkat, ezzel pontosítva az ajánlást. Nagyon egyszerű példa lehet egy szálloda, illetve nyaralóhely kereső portál, ahol a felhasználó kiválaszthatja a számára megfelelő helyet. Ebben az esetben az ajánlórendszer figyelembe veheti az évszakot, hiszen egészen más ajánlatokat lehet adni télen, mint nyáron. Egy valós példa a Sourcetone (<http://www.sourcetone.com/>) nevű interaktív rádió, ami figyelembe veszi a felhasználó aktuális kedélyállapotát (kontextus információ), amit a meghallgatásra kiválasztott zeneszámokból próbál kikövetkeztetni.

A megoldás figyelembe veszi a kontextus információkat is, amik lehetnek (többek között):

- idő alapú: télen sí-, nyáron búvárfelszerelés,
- hangulat alapú: ha romantikus zenét hallgatott, ajánljon romantikus filmet is,
- készülék alapú: nagyobb képernyőre filmet, kisebb eszközre zenét,
- lokáció, vagy esemény alapú

Ajánlási technikák I.

▶ Tartalom alapú szűrés

- ▶ Olyan elemeket javasol, amik hasonlóak a felhasználó eddigi választásaihoz.
- ▶ A hasonlóság meghatározása az elemek meta adatai alapján történik.

▶ Kollaboratív szűrés

- ▶ Olyan elemeket javasol, amiket az aktuális felhasználóhoz hasonló felhasználók szeretnek.
- ▶ A felhasználók közötti hasonlóság mérése az eddigi interakciók alapján történik.

▶ Demográfiai

- ▶ Az ajánlás demográfiai profil alapján készül.
- ▶ Informatikát tanuló fiataloknak számítógéphez kapcsolódó elemeket javasol.

Ajánlási technikák II.

▶ Tudás alapú

- ▶ A felhasználóról és az elemekről tudás áll rendelkezésre, az ajánlás ezek felhasználásával történik.
- ▶ Ha a felhasználó fényképezőgépet keres, a rendszer a olyat javasol, ami a pénztárcájának és a tudásának megfelelő.

▶ Közösség alapú

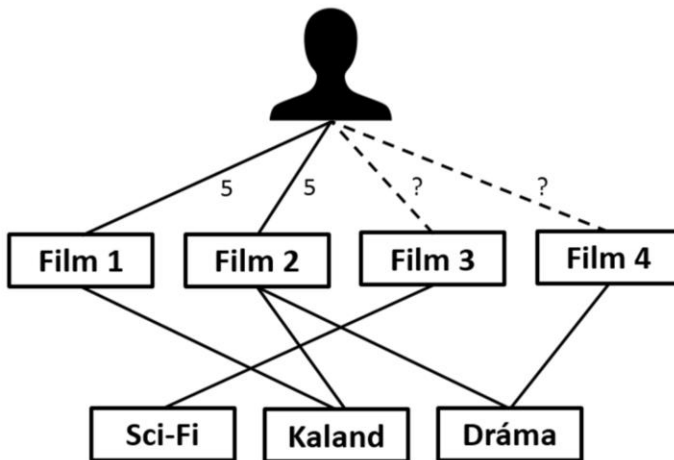
- ▶ Az ajánlás alapja az adott felhasználó baráti köre.
- ▶ Ismerősöktől szívesebben fogadunk el ajánlást (közösségi oldalak alapján).

▶ Hibrid

- ▶ Az fentiek kombinációja.
- ▶ Az előnyök kihasználásával és a hátrányok javításával.

13

Tartalom alapú szűrés



14

Vannak elemek (filmek), azokról rendelkezünk bizonyos információval.
Feladat annak megbecslése, hogy a még nem értékelt filmeket mennyire szeretné a felhasználó.

Tartalom alapú szűrés

▶ Elemek

- ▶ Elem modellt épít.
- ▶ Meta adatok alapján (pl. műfaj).

▶ Felhasználók

- ▶ Felhasználói profilt épít.
- ▶ A tranzakciós történet alapján (pl. Film 1-et és Film 2-öt 5-ösre értékelt).

▶ Ajánlás

- ▶ A felhasználói profilt hasonlítja össze az elemreprezentációkkal.
- ▶ A hasonló elemeket javasolja.
 - ▶ Hasonló azokhoz, amiket eddig magasra értékelt.

15

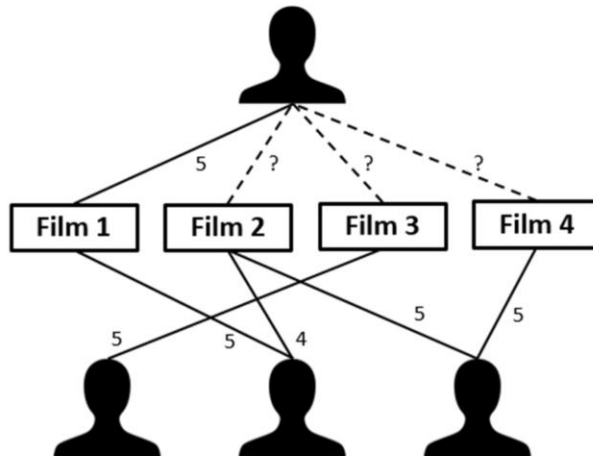
Tartalom alapú szűrés / Előnyök és hátrányok

- + Felhasználó-függetlenség
 - ▶ A profilok egymástól függetlenül épülnek, másik felhasználóé nem befolyásolja.
- + Nincs elem cold start probléma
- Túlspecializált
 - ▶ Nem képes váratlant ajánlani, rátanult a felhasználó preferenciájára.
- Tartalom analízis
 - ▶ A leírók minősége, száma nagyban befolyásolja az összehasonlítást.
- Felhasználói cold start
 - ▶ Új felhasználónak nem tudja pontosan elkészíteni a profilját.

16

Cold start probléma: az ajánlás során gyakran felmerülő probléma, amelynek lényege a következő. Bizonyos esetekben, amikor új felhasználó, vagy elem érkezik, a rendszer nem tud vele mit kezdeni, mert nincs róla semmilyen előzetes információja. Nem tud neki mit ajánlani, mert nem ismeri a preferenciáját.

Kollaboratív szűrés



17

Olyan elemeket javasol, amiket olyan felhasználók értékelték magas pontszámmal, akiknek hasonló ízlése van az adott felhasználóhoz.

A felhasználók közötti összehasonlítás a múltbeli tranzakciók alapján történik.

Cél megjósolni, hogy a felhasználó milyenre értékelné az ismeretlen elemeket.

A tranzakciós adatok használatának következménye, hogy az ajánlási módszer domain független, tehát nem csak egy típusba tartozó elemek ajánlását valósítja meg. Például ha az adott felhasználóhoz hasonló felhasználók preferenciájába könyvek, filmek, és zeneszámok is tartoznak, akkor a rendszer az aktuális felhasználó számára is ezekből a domainekből fog ajánlani.

Kollaboratív szűrés

- ▶ **Globális átlag**
 - ▶ Az összes értékelés átlaga.
- ▶ **Felhasználó átlag**
 - ▶ Az adott felhasználó értékeléseinek az átlaga.
- ▶ **Elem átlag**
 - ▶ Az adott elemhez tartozó értékelések átlaga.
- ▶ **Szomszédok módszere**
- ▶ **Mátrix faktorizáció**

18

Kollaboratív szűrés / Szomszédok módszere

▶ Elem alapú

- ▶ A javaslatot az adott elem szomszédjai alapján hozza meg.
- ▶ Kiválaszt k darab hasonló elemet, amit már értékelt a felhasználó.
- ▶ Ezekből az értékelésekből becsli meg az új elem relevanciáját.

▶ Felhasználó alapú

- ▶ A javaslatot az adott felhasználó szomszédjai alapján határozza meg.
- ▶ Kiválaszt k darab hasonló felhasználót, akik már értékelték az adott elemet.
- ▶ Ezekből az értékelésekből becsli meg az új elem relevanciáját.

19

Elem alapú: Például, ha az – eddig még nem értékelt – Mátrix harmadik részére próbál a rendszer becslést adni u_a , akkor megvizsgálja annak szomszédjait: valószínűleg ide sorolhatóak az előző részei, a tudományos fantasztikus műfajba sorolható, illetve a Keanu Reeves szereplésében játszódó filmek. Ezek közül a filmek közül azok jelentik a becslés alapját, amiket a felhasználó már előzőleg értékelt.

Felhasználó alapú: Az előző példa tükrében, olyan felhasználókat keres a rendszer, akiknek hasonló az ízlése és már értékelték a Mátrix harmadik részét, majd ezek közül kiválasztja azokat, akik hasonlóan értékelték az eddig megnézett filmeket, mint az aktuális felhasználó.

Kollaboratív szűrés / Mátrix faktorizáció

M	Film 1	Film 2	Film 3	Film 4	U	
Felh. 1	5				1	1
Felh. 2			5		1	1
Felh. 3	5	4			1	1
Felh. 4		5		5	1	1

V	Film 1	Film 2	Film 3	Film 4
	1	1	1	1
	1	1	1	1

Cél: $U \times V = M$

20

M hiányzó értékeit U és V segédmátrixokkal próbálja meghatározni. U és V kezdetben legyen feltöltve (pl. csupa 1). M ismert értékeinek segítségével U és V mátrixot módosítja úgy, hogy $U \times V = M$ teljesüljön.

Kollaboratív szűrés / Mátrix faktorizáció

M	Film 1	Film 2	Film 3	Film 4	U	
Felh. 1	5				2	1
Felh. 2			5		1	2
Felh. 3	5	4			2	1
Felh. 4		5		5	5	0

V	Film 1	Film 2	Film 3	Film 4
	2	1	4	1
	1	2	1	0

$$4 \cdot 1 + 1 \cdot 2 = 6$$

21

Ha az adott elemre $UxV=M$ nem teljesül módosítja U vagy V értékét kicsivel.

Kollaboratív szűrés / Matrik faktorizáció

M	Film 1	Film 2	Film 3	Film 4	U	
Felh. 1	5				2	1
Felh. 2			5		1	2
Felh. 3	5	4			2	1
Felh. 4		5		5	5	0

V	Film 1	Film 2	Film 3	Film 4
	2	1	3	1
	1	2	1	0

$$3 \cdot 2 + 1 \cdot 1 = 5$$

22

Miután a segédmatrikokat M ismert értékei alapján kitöltöttük, feladat M matrik becslése U és V segítségével, $UxV=M$ alapján kitölti.

Kollaboratív szűrés / Előnyök és hátrányok

- + Hatékonyabb mint a tartalom alapú megoldás
- + Domain független
 - ▶ Egy hasonló felhasználó preferenciájába könyv és film is tartozik, azokat is javasolja.
- Elem cold start probléma
 - ▶ Értékelés nélküli elemmel nem tud mit kezdeni a rendszer.
- Bemelegedési fázis
 - ▶ Csak akkor hatékony, ha már sok tranzakció volt.
- Nehezen magyarázható
 - ▶ Nehéz visszajelezni a felhasználónak, hogy miért történt az ajánlás.

23

Összefoglalás

- ▶ Tartalom alapú ajánlás
 - ▶ Rotten Tomatos, IMDB
- ▶ Kollaboratív szűrés alapú ajánlás
 - ▶ Twitter, Facebook, Last.fm, LinkedIn, Amazon
- ▶ Új irány: hibrid ajánlás
 - ▶ ImpressTV