

# Üzleti Elemzések Módszertana

## 9. Előadás: Ajánló rendszerek

Kuknyó Dániel  
Budapesti Gazdasági Egyetem

2023/24  
2.félév

1 Bevezetés

2 Kollaboratív szűrők

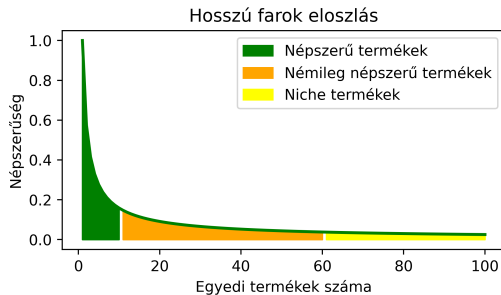
1 Bevezetés

2 Kollaboratív szűrők

# Hosszú farok eloszlás

A webes vásárlás elterjedése előtt az volt a jellemző, hogy kevés termék generálta a forgalom legnagyobb hányadát. Mivel az üzlethelyiségben a férőhely limitált volt, a kevesek által keresett termékek nem kaptak helyet a polcon.

Az internetes kereskedelem elterjedése helyet adott az ún. *niche*, vagyis szűk csoportok számára népszerű termékeknek, amelyek specifikus felhasználásukkal vonzzák be a célközönséget

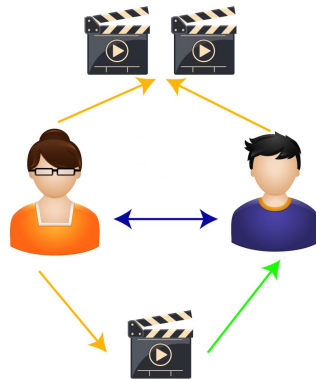


# Ajánló rendszerek

Olyan technikák vagy rendszerek, amelyek **valamilyen terméket, szolgáltatást vagy entitást kötnek össze más termékekkel, javaslatokkal vagy entitásokkal a rendelkezésükre álló információ alapján.**

Az ajánló rendszerek célja objektumok közötti leképezések felderítése, mint:

- Filmek
- Termékek
- Könyvek
- Média

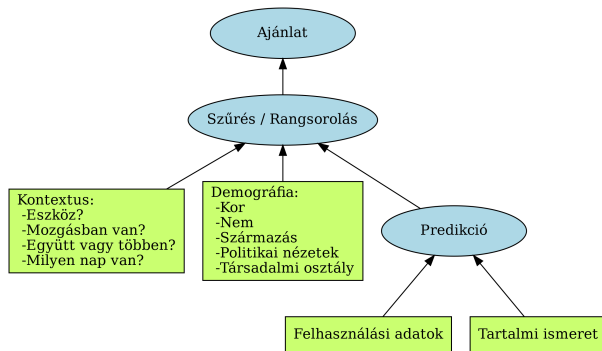


# Ajánlat előállításának szintjei

A rendszer célja, hogy ajánlásokat tegyen a rendelkezésre álló információ alapján.

Különböző rendszertípusoknak különböző igénye van az adatok forrásával, milyenségével és rendelkezésre állásával szemben.

A rendszereket lehetséges célnak megfelelően optimalizálni a népszerűség vagy személyre szabottságnak megfelelően.

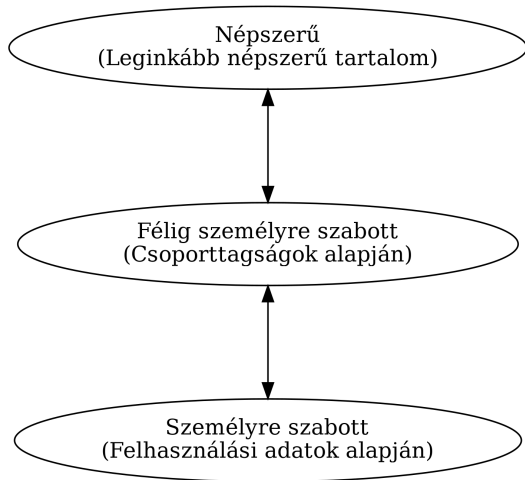


# Ajánlatok személyre szabottsága

Egy ajánló rendszer által adott ajánlat a személyre szabottól a népszerűig terjedhet.

A személyre szabott ajánlatok a felhasználó egyedi igényét célozzák, és ezért több felhasználási adatot igényelnek.

A népszerűségeen alapuló ajánlatoknak nincs szüksége felhasználási adatokra, de nem képes az egyén ízlésének megfelelő ajánlatot előállítani.



# A predikciós probléma

Adott egy  $m$  felhasználóból és  $n$  termékből álló  $X$  mátrix.  $X[u, i]$  azt jelöli, hogy  $u$  felhasználó hogyan értékelte  $i$  terméket.

A feladat megbecsülni a mátrix hiányzó értékeit a termékekről és felhasználókról rendelkezésre álló információ alapján.

Ez a mátrix leggyakrabban egy ritka mátrix: nagyon sok az ismeretlen érték.

	$i_1$	$i_2$	$i_3$	$i_4$	$i_5$	$i_6$
$u_1$	4	?	3	?	5	?
$u_2$	?	2	?	?	4	1
$u_3$	?	?	1	?	2	5
$u_4$	?	?	3	?	?	1
$u_5$	1	4	?	?	2	5
$u_6$	5	?	2	1	?	4
$u_7$	?	2	3	?	4	5

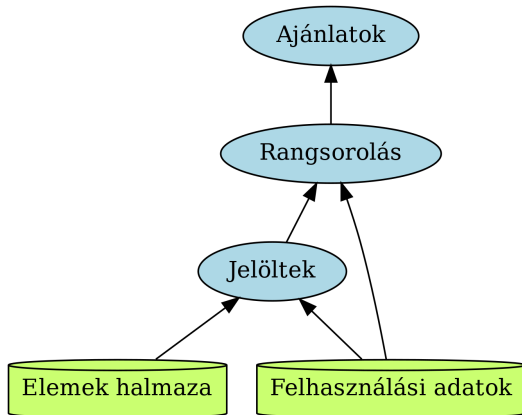


# A rangsorolási probléma

A rangsorolási probléma a predikciós problémának egy intuitívabb megfogalmazása.

Ha adott  $n$  elem halmaza, a rangsorolás célja megkülönböztetni a leginkább javasolható  $k$  elemet, amit ajánlhat a felhasználónak valamilyen rendezési kritérium alapján.

A predikciós probléma gyakran rangsorolási problémához vezet vissza.



1 Bevezetés

2 Kollaboratív szűrők

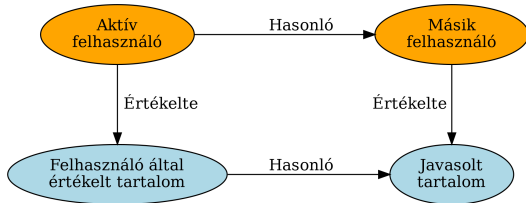
# Kollaboratív szűrők

## Kollaboratív szűrő

Olyan ajánló rendszer, amely a közösség által adott értékelésekből, felhasználási metrikákból állít össze javaslatokat.

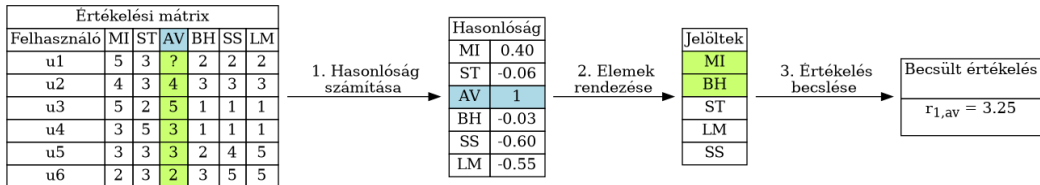
Két típusa létezik:

- **Felhasználó alapú:** Az adott felhasználónak a hozzá hasonló felhasználók preferenciái alapján ajánl termékeket.
- **Termék alapú:** Olyan elemeket ajánl a felhasználónak, amelyek hasonlóak az általa preferált elemekhez.



# Termék alapú kollaboratív szűrő eljárása

A példában a cél  $u_1$  felhasználó értékelésének megbecsülése az AV filmhez. Ehhez tartozóan az eljárás először kiszámítja az AV film hasonlóságát az összes többivel, majd ezeket rangsorolja és a leghasonlóbb elemek értékelései alapján megadja annak becsült értékelését.



# Két egyed hasonlóságának kiszámítása

## Elemek hasonlósága

$$\text{sim}(i_1, i_2) = \frac{\sum_u (nr_{i_1, u} \cdot nr_{i_2, u})}{\sqrt{\sum_u nr_{i_1, u}^2} \cdot \sqrt{\sum_u nr_{i_2, u}^2}}$$

Ahol:

- $r_{i,u}$ :  $u$  felhasználó értékelése  $i$  elemre
- $\bar{r}_u$ :  $u$  felhasználó átlagos értékelése
- $nr_{i,u} = r_{i,u} - \bar{r}_u$ : Normalizálási tényező

## Felhasználók hasonlósága

$$\text{sim}(u_1, u_2) = \frac{\sum_i (nr_{i, u_1} \cdot nr_{i, u_2})}{\sqrt{\sum_i (nr_{i, u_1})^2} \cdot \sqrt{\sum_i (nr_{i, u_2})^2}}$$

Ahol:

- $r_{i,u}$ :  $u$  felhasználó értékelése  $i$  elemre
- $\bar{r}_u$ : az  $u$  felhasználó átlagos értékelése
- $nr_{i,u} = r_{i,u} - \bar{r}_u$ : az értékelések normalizálása, a felhasználó átlagától való eltérés

# Hasonlóság kiszámítása minden mintaegyedre

Az ajánlásokhoz érdemes a hasonlóságokat előzetesen kiszámolni, hogy csökkentse a teljes rendszer erőforrásigényét. Ehhez tartozóan a hasonlósági mátrix létrehozásának algoritmus:

---

**Algoritmus 1:** Termék-termék kollaboratív szűrő

---

```
for Termék  $i_1$  a katalógusban do
  for Felhasználó  $u$  aki fogyasztotta  $i_1$  terméket do
    for Termék  $i_2$  amit  $u$  fogyasztott do
      | Termékpáros  $(i_1, i_2)$  rögzítése
    end
    for Minden  $i_2$  termékre do
      |  $sim(i_1, i_2)$  kiszámítása
    end
  end
end
```

---

## Példa: hasonlósági tábla összeállítása

Adott az alábbi értékelési mátrix 6 filmmel és 6 felhasználóval. A cél megbecsülni a táblázat hiányzó értékeit. Első lépésben el kell menteni, melyik filmeket látták még azok a felhasználók, amelyek egy adott filmet megnézték:

*MIB* :

[*ST*, *B*, *SS*, *LM*, *AV*]

*ST* :

[*MIB*, *B*, *SS*, *LM*, *AV*]

*B* :

[*MIB*, *ST*, *SS*, *LM*, *AV*]

*SS* :






[*MIB*, *ST*, *B*, *LM*, *AV*]

*LM* :

[*MIB*, *ST*, *SS*, *B*, *AV*]






*AV* :

[*MIB*, *ST*, *B*, *SS*, *LM*]

						
Név	MIB	ST	AV	B	SS	LM
Sara	2.20	0.20	?	-0.80	-0.80	-0.80
Jesper	0.60	-0.40	0.60	?	-0.40	-0.40
Therese	2.33	-0.67	2.33	-0.67	-1.67	-1.67
Helle	0.40	2.40	0.40	?	-1.60	-1.60
Pietro	-0.33	-0.33	-0.33	-1.33	0.67	1.67
Ekaterina	-1.33	-0.33	-1.33	0.33	1.67	1.67

## Normalizált értékelési tábla







Az értékelési tábla és az adott felhasználó értékelései szerint felírható a normalizált értékelési tábla. A pozitív értékelések az adott felhasználó átlagos értékelésénél jobbnak számítanak, míg a negatívak rosszabbnak.

						
Név	MIB	ST	AV	B	SS	LM
Sara	2.20	0.20	?	-0.80	-0.80	-0.80
Jesper	0.60	-0.40	0.60	?	-0.40	-0.40
Therese	2.33	-0.67	2.33	-0.67	-1.67	-1.67
Helle	0.40	2.40	0.40	?	-1.60	-1.60
Pietro	-0.33	-0.33	-0.33	-1.33	0.67	1.67
Ekaterina	-1.33	-0.33	-1.33	-0.33	1.67	1.67



# Korreláció számítása

A normalizált értékelésekre korrelációs együtthatót számítva előáll a normalizált termék-termék korrelációs mátrixot. Az 1 érték jelenti a tökéletes hasonlóságot, -1 érték a tökéletes különbözőséget, 0 pedig a közömbös kapcsolatot.

						
Név	MIB	ST	AV	B	SS	LM
MIB	1	0.63	1	-0.21	0.88	-0.83
ST	0.63	1	0.35	-0.47	-0.64	-0.62
AV	1	0.35	1	0.01	-0.89	-0.83
B	-0.21	-0.47	0.01	1	-0.23	-0.32
SS	0.88	-0.64	-0.89	-0.23	1	0.96
LM	-0.83	-0.62	-0.83	-0.32	0.96	1