

Üzleti Elemzések Módszertana

9. Előadás: Ajánló rendszerek

Kuknyó Dániel

Budapesti Gazdasági Egyetem

2023/24

2.félév

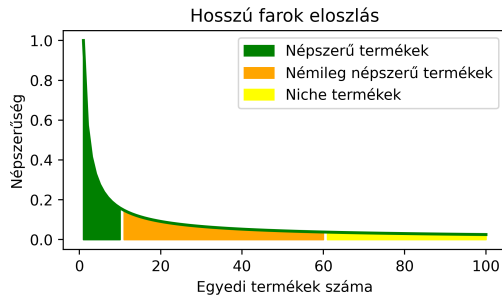
- 1 Bevezetés
- 2 Kollaboratív szűrők
- 3 Tartalom alapú rendszerek
- 4 Tudásalapú rendszerek
- 5 Hibrid ajánlók

- 1 Bevezetés
- 2 Kollaboratív szűrők
- 3 Tartalom alapú rendszerek
- 4 Tudásalapú rendszerek
- 5 Hibrid ajánlók

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

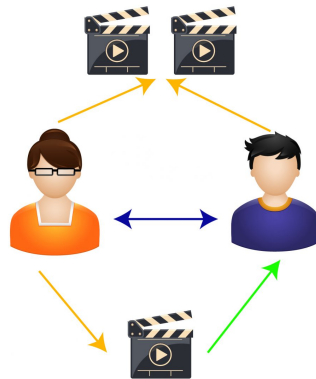


Ajánló rendszerek

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

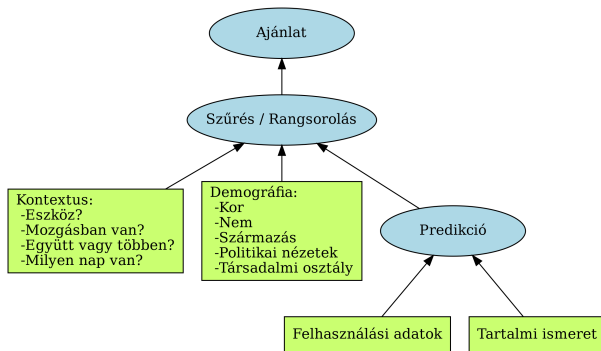


Ajánlat előállításának szintjei

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.

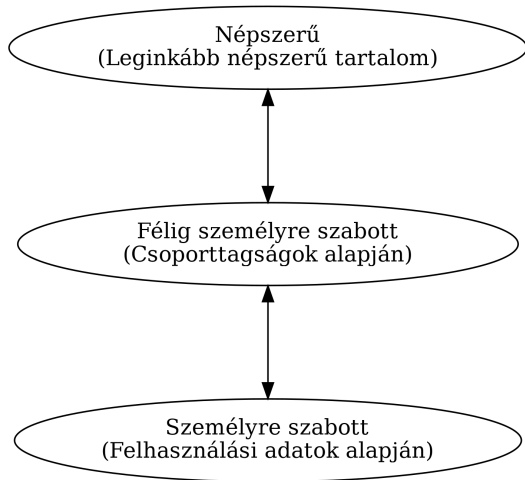


Ajánlatok személyre szabottsága

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.

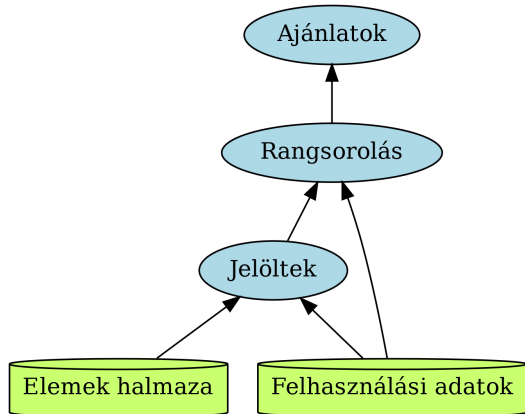
	i_1	i_2	i_3	i_4	i_5	i_6
u_1	4	?	3	?	5	?
u_2	?	2	?	?	4	1
u_3	?	?	1	?	2	5
u_4	?	?	3	?	?	1
u_5	1	4	?	?	2	5
u_6	5	?	2	1	?	4
u_7	?	2	3	?	4	5

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
- 3 Tartalom alapú rendszerek
- 4 Tudásalapú rendszerek
- 5 Hibrid ajánlók

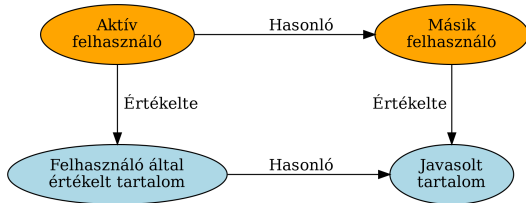
Kollaboratív szűrők

Kollaboratív szűrő

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

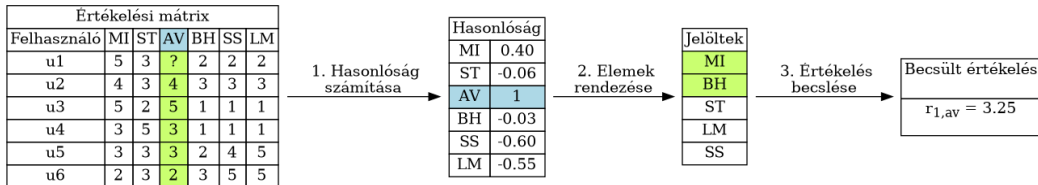
Két típusa létezik:

- **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:

						
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

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

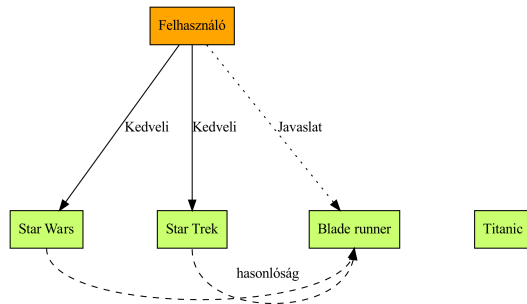
- 1 Bevezetés
- 2 Kollaboratív szűrők
- 3 Tartalom alapú rendszerek**
- 4 Tudásalapú rendszerek
- 5 Hibrid ajánlók

Tartalom alapú rendszerek

A kollaboratív szűrőkkel ellentétben a tartalom alapú szűrőknek nincs szüksége információra múltbeli vásárlásokról és információról.

Ehelyett a javaslatokat a **felhasználók profiljai és a termékek metaadatai alapján állítják össze.**

A tartalom alapú rendszerek viszont nem használják ki a közösség adta lehetőségeket.



Tartalom alapú rendszerek csővezetéke

A tartalom alapú rendszerek esetén a rendszer úgy tesz ajánlásokat, hogy a felhasználó által fogyasztott tartalmakhoz hasonló elemeket mutat neki ajánlásként.

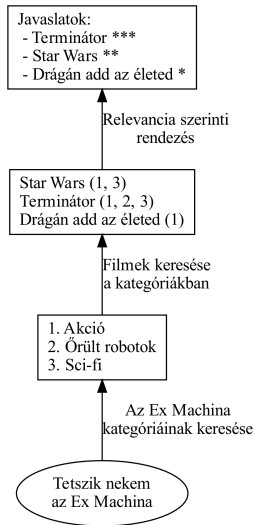
Az objektumokhoz tartozó két fontos tartalmi leíró elem a **tag** és a **fact**.

Tag

A Web2.0 idején megjelent weboldal tartalomra utaló kulcsszavak.

Fact

Tényszerűen írja le az adott elemet.

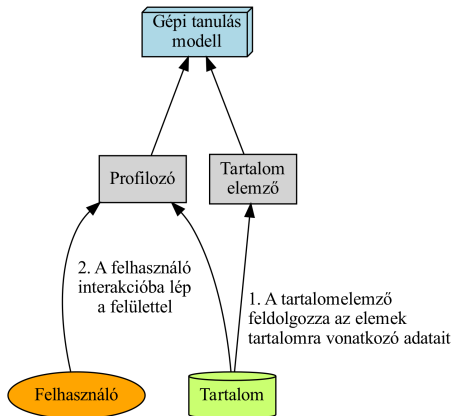


Tartalomalapú rendszerek komponensei

Tartalomelemző

A tartalom alapján készít modelleket az elérhető elemekről.

Tartalomalapú rendszerek tanítása

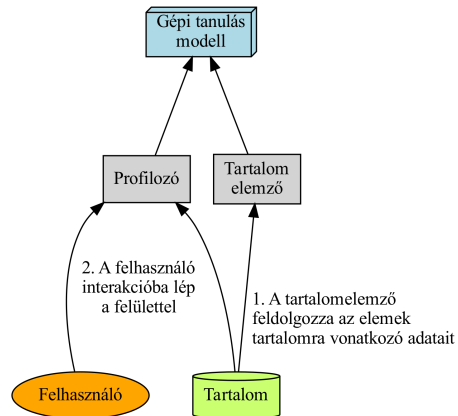


Tartalomalapú rendszerek komponensei

Felhasználó profilozó

Felhasználói profilokat készít. Ez gyakran egy egyszerű lista azokról az elemekről, amiket a felhasználó fogyasztott.

Tartalomalapú rendszerek tanítása

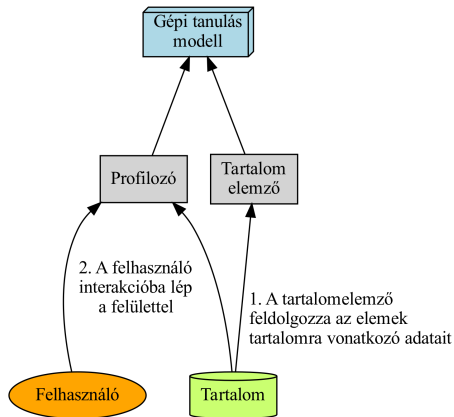


Tartalomalapú rendszerek komponensei

Elem visszakereső

Célja megkeresni a releváns tartalmakat azáltal, hogy a felhasználói profilokat összehasonlítja az elemek modelljeivel.

Tartalomalapú rendszerek tanítása

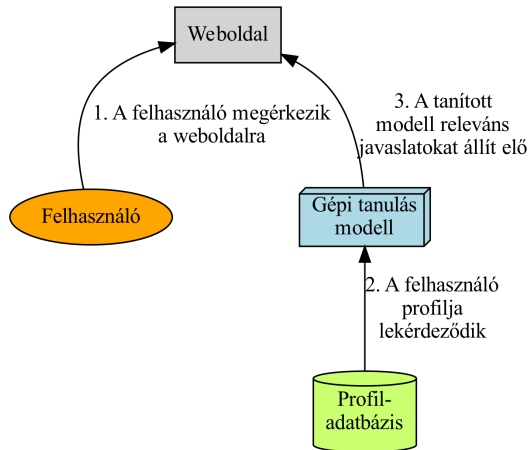


Tartalomalapú rendszerek komponensei

Elem visszakereső

Célja megkeresni a releváns tartalmakat azáltal, hogy a felhasználói profilokat összehasonlítja az elemek modelljeivel.

Predikció tartalomalapú rendszerek esetén



- 1 Bevezetés
- 2 Kollaboratív szűrők
- 3 Tartalom alapú rendszerek
- 4 Tudásalapú rendszerek**
- 5 Hibrid ajánlók

Tudásalapú rendszerek

Azon tételek esetén használatos, amelyeket a felhasználók ritkán vásárolnak, és ezért nagyon specifikus igényeik lehetnek az elemekkel kapcsolatban. Pl.: ingatlanok, magas értékű elektromos cikkek.

A kereséshez a rendszerben a felhasználónak meg kell adnia az **elvárt paramétereket**, és a **rendszer visszatéríti azokat az egyedeket, amelyekre jellemzőek ezek a feltételek**. Hátránya, hogy sosem fog váratlan, újszerű eredményeket visszatéríteni.

Operációs rendszer	▼
Processzor típusa	▼
Processzor órajel	▼
Cache mérete	▼
Memória mérete	▼
Memória típusa	▼
Kijelző mérete	▼

- 1 Bevezetés
- 2 Kollaboratív szűrők
- 3 Tartalom alapú rendszerek
- 4 Tudásalapú rendszerek
- 5 Hibrid ajánlók

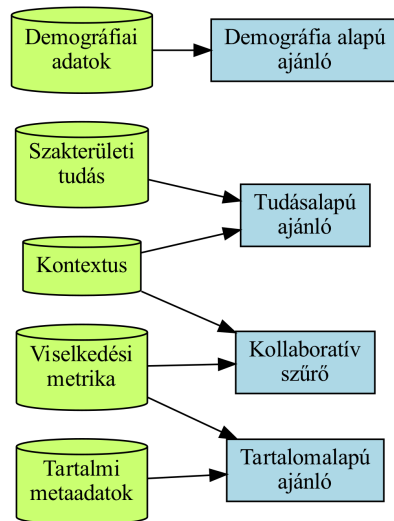
Az ajánló rendszerek megközelítései

Minden ajánló rendszer típusnak megvannak a maga előnyei és hátrányai.

A hibrid rendszerek célja **adott komponensek pozitív hozadékainak megtartása, míg a negatív tulajdonságainak ellensúlyozása** más rendszerek előnyei által.

A hibrid rendszerek három fő típusa:

- Monolitikus
- Együttes
- Kevert



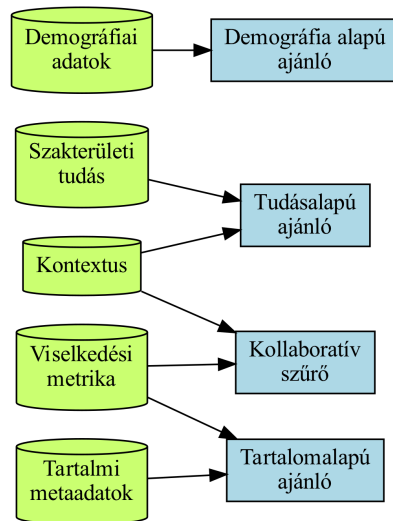
Az ajánló rendszerek megközelítései

Hidegindítási probléma

A hidegindítási probléma az ajánló rendszerekben akkor jelentkezik, amikor az új felhasználók vagy új elemek **nem rendelkeznek elegendő interakciós adattal** ahhoz, hogy az ajánló algoritmusok hatékonyan dolgozhassanak.

Hidegindítási problémától szenvedő ajánló típusok:

- Kollaboratív szűrők
- Demográfiaalapú rendszerek

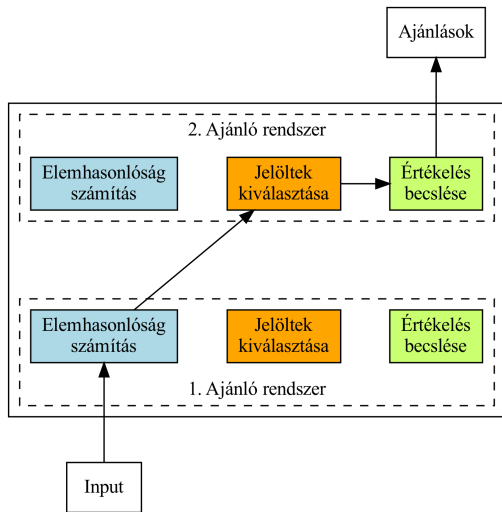


Hibrid ajánlók: monolitikus rendszerek

Egy ajánló rendszer általánosságban több komponenst tartalmaz adott feladatok elvégzésére, mint a hasonlóság számító, elemkiválasztó.

Egy monolitikus ajánló **különböző rendszerek komponenseit használja fel egy csővezetékbe építve.**

Például: tartalomalapú adatok keverése viselkedéshez köthető javaslatokkal.

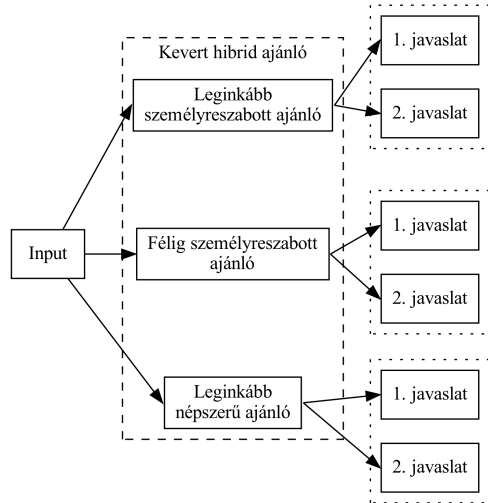


Hibrid ajánlók: kevert rendszerek

A kevert hibrid rendszerek **szigetszerűen működő ajánló rendszerek predikcióinak unióját térítik vissza.**

Az ajánló rendszerek szigetei terjedhetnek például a leginkább személyreszabottól a legnépszerűbbig. Gyakran a személyreszabott rendszerek 1-2 predikciót adnak, míg a népszerű elemeket célzó rendszerek sokkal többet.

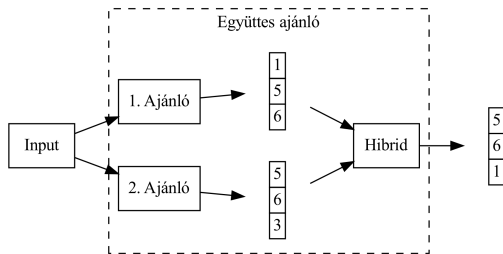
A rendszerek által adott predikciókat normalizálva lehetséges a fontossági lista előállítás.



Hibrid ajánlók: együttes rendszerek

Ahogy az együttes tanulás esetében, úgy az **együttes ajánló rendszerek is több különböző rendszer predikcióit aggregálják a végleges predikció létrehozásához**. Az aggregáció történhet különböző módszerekkel, mint a szavazás, súlyozás kapcsolás.

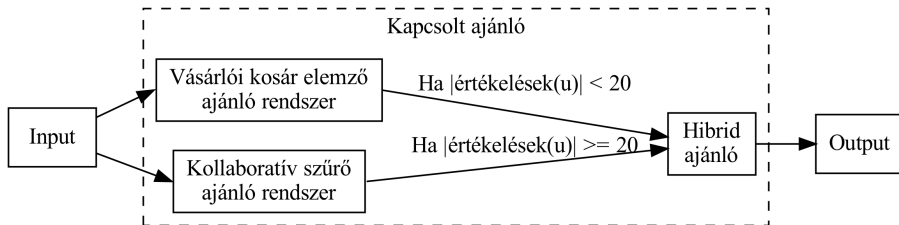
A példában R_1 predikciói $[1, 5, 6]$, R_2 predikciói pedig $[6, 5, 3]$. Ebben az esetben a végső predikció $[5, 6, 1]$ lesz az aggregálás után.



Hibrid ajánlók: kapcsolt rendszerek

A kapcsolt ajánló rendszerek valamilyen **feltétel szerint választanak adott modellek predikciói közül**. Ebben a példában a feltétel, hogy mennyi terméket értékelt a felhasználó.

Ennek az intuíciója, hogy aki keveset értékelt, annak a vásárlói kosara több releváns információt tartalmazhat.



Hibrid ajánlók: súlyozott rendszerek

Különböző rendszerek adottságait **súlyozás** segítségével is fel lehet használni együttesen.

A tartalomalapú rendszer nem tesz különbséget jó és rossz minőség között, míg a kollaboratív szűrés nem tesz különbséget fontosságban.

A súlyozott rendszerek a komponenseik predikcióit valamilyen súlyozás szerint veszik figyelembe, ennek alapján áll elő a végső predikció.

