Ajánló rendszerek – áttekintés

Hegedűs István, Ormándi Róbert

Ajánlási feladat definíciója

- Adott:

Ajánlási feladat definíciója

- Adott:
  - U - felhasználók (users),

Ajánlási feladat definíciója

- Adott:
  - U - felhasználók (users),
  - I - termékek (items),

Ajánlási feladat definíciója

- Adott:
  - U - felhasználók (users),
  - I - termékek (items),
  - $p : U \rightarrow profileSpace$ - felhasználói profilolok,

Ajánlási feladat definíciója

- Adott:
  - U - felhasználók (users),
  - I - termékek (items),
  - $p : U \rightarrow profileSpace$ - felhasználói profilolok,
  - $c : I \rightarrow characteristicSpace$ - termék leírások,

Ajánlási feladat definíciója

- Adott:
  - U - felhasználók (users),
  - I - termékek (items),
  - $p : U \rightarrow profileSpace$ - felhasználói profilolok,
  - $c : I \rightarrow characteristicSpace$ - termék leírások,
  - egy $f : U \times I \rightarrow R$ leképezésből néhány példa (R halmaz, teljes rendezéssel).

Ajánlási feladat definíciója

- Adott:
  - U - felhasználók (users),
  - I - termékek (items),
  - $p : U \rightarrow profileSpace$ - felhasználói profilolok,
  - $c : I \rightarrow characteristicSpace$ - termék leírások,
  - egy $f : U \times I \rightarrow R$ leképezésből néhány példa (R halmaz, teljes rendezéssel).
- Feladat:

Ajánlási feladat definíciója

- Adott:
  - U - felhasználók (users),
  - I - termékek (items),
  - $p : U \rightarrow profileSpace$ - felhasználói profilolok,
  - $c : I \rightarrow characteristicSpace$ - termék leírások,
  - egy $f : U \times I \rightarrow R$ leképezésből néhány példa (R halmaz, teljes rendezéssel).
- Feladat:
  - Határozzuk meg azt az $\hat{f} : U \times I \rightarrow R$ leképezést, amely

Ajánlási feladat definíciója

- Adott:
  - U - felhasználók (users),
  - I - termékek (items),
  - $p : U \rightarrow profileSpace$ - felhasználói profilolok,
  - $c : I \rightarrow characteristicSpace$ - termék leírások,
  - egy $f : U \times I \rightarrow R$ leképezésből néhány példa (R halmaz, teljes rendezéssel).
- Feladat:
  - Határozzuk meg azt az $\hat{f} : U \times I \rightarrow R$ leképezést, amely
    - a lehető legjobban közelíti f-et, és

Ajánlási feladat definíciója

- Adott:
  - U - felhasználók (users),
  - I - termékek (items),
  - $p : U \rightarrow profileSpace$ - felhasználói profilolok,
  - $c : I \rightarrow characteristicSpace$ - termék leírások,
  - egy $f : U \times I \rightarrow R$ leképezésből néhány példa (R halmaz, teljes rendezéssel).
- Feladat:
  - Határozzuk meg azt az $\hat{f} : U \times I \rightarrow R$ leképezést, amely
    - a lehető legjobban közelíti f-et, és
    - teljesen definiált a teljes $U \times I$ téren.

Megjegyzések

- Ajánlási feladat megoldása: algoritmus, amely meghatározza $\hat{f}$-et.

Megjegyzések

- Ajánlási feladat megoldása: algoritmus, amely meghatározza $\hat{f}$-et.
- Ajánló rendszer: rendszer, ami használja $\hat{f}$-et, pl.:

Megjegyzések

- Ajánlási feladat megoldása: algoritmus, amely meghatározza $\hat{f}$-et.
- Ajánló rendszer: rendszer, ami használja $\hat{f}$-et, pl.:
  - u felhasználónak megvételjre ajánlja az i_u terméket, melyre

Megjegyzések

- Ajánlási feladat megoldása: algoritmus, amely meghatározza $\hat{f}$-et.
- Ajánló rendszer: rendszer, ami használja $\hat{f}$-et, pl.:
  - u felhasználónak megvételjre ajánlja az i_u terméket, melyre
  - $i_u = \text{argmax}_{i \in I, f(i,u) \text{definiálatlan}} \hat{f}(u, i)$.

Megjegyzések

- Ajánlási feladat megoldása: algoritmus, amely meghatározza $\hat{f}$-et.
- Ajánló rendszer: rendszer, ami használja $\hat{f}$-et, pl.:
  - u felhasználónak megvételjre ajánlja az i_u terméket, melyre
  - $i_u = \text{argmax}_{i \in I, f(i,u) \text{definiálatlan}} \hat{f}(u, i)$.
- Az f-ből ismert értékelések aránya, a teljes $U \times I$ tér méretéhez képest nagyon kicsi.

Algoritmusok osztályozása

- Felhasznált információ alapján:

Algoritmusok osztályozása

- Felhasznált információ alapján:
  - content based methods: a cél felhasználó korábbi értékelései alapján történik a predikció

Algoritmusok osztályozása

- Felhasznált információ alapján:
  - content based methods: a cél felhasználó korábbi értékelései alapján történik a predikció
  - collaborative filtering: a predikció inputjához hasonló adatok is felhasználásra kerülnek a predikció során

Algoritmusok osztályozása

- Felhasznált információ alapján:
  - content based methods: a cél felhasználó korábbi értékelései alapján történik a predikció
  - collaborative filtering: a predikció inputjához hasonló adatok is felhasználásra kerülnek a predikció során
    - user-based: a cél felhasználóhoz hasonló izlésű felhasználók szavazatainak felhasználásával történik a predikció

Algoritmusok osztályozása

- Felhasznált információ alapján:
  - content based methods: a cél felhasználó korábbi értékelései alapján történik a predikció
  - collaborative filtering: a predikció inputjához hasonló adatok is felhasználásra kerülnek a predikció során
    - user-based: a cél felhasználóhoz hasonló izlésű felhasználók szavazatainak felhasználásával történik a predikció
    - item-based: a cél termékhez hasonló termékekre tett értékelések használatával készül a predikció

Algoritmusok osztályozása

- Felhasznált információ alapján:
  - content based methods: a cél felhasználó korábbi értékelései alapján történik a predikció
  - collaborative filtering: a predikció inputjához hasonló adatok is felhasználásra kerülnek a predikció során
    - user-based: a cél felhasználóhoz hasonló izlésű felhasználók szavazatainak felhasználásával történik a predikció
    - item-based: a cél termékhez hasonló termékekre tett értékelések használatával készül a predikció
  - hybrid methods: az előző két módszer ötvözete

Algoritmusok osztályozása

- Az algoritmusban használt megközelítés alapján:

Algoritmusok osztályozása

- Az algoritmusban használt megközelítés alapján:
  - hasonlóság alapú megközelítések,

Algoritmusok osztályozása

- Az algoritmusban használt megközelítés alapján:
  - hasonlóság alapú megközelítések,
  - mátrix faktorizáción alapuló módszerek,

Algoritmusok osztályozása

- Az algoritmusban használt megközelítés alapján:
  - hasonlóság alapú megközelítések,
  - mátrix faktorizáción alapuló módszerek,
  - klaszterezést hasznláló módszerek,

Algoritmusok osztályozása

- Az algoritmusban használt megközelítés alapján:
  - hasonlóság alapú megközelítések,
  - mátrix faktorizáción alapuló módszerek,
  - klaszterezést használó módszerek,
  - gépi tanulási modellt használó eljárások,

Algoritmusok osztályozása

- Az algoritmusban használt megközelítés alapján:
  - hasonlóság alapú megközelítések,
  - mátrix faktorizáción alapuló módszerek,
  - klaszterezést használó módszerek,
  - gépi tanulási modellt használó eljárások,
  - valószínűségi modell alapú módszerek,

Algoritmusok osztályozása

- Az algoritmusban használt megközelítés alapján:
  - hasonlóság alapú megközelítések,
  - mátrix faktorización alapuló módszerek,
  - klaszterezést használó módszerek,
  - gépi tanulási modellt használó eljárások,
  - valószínűségi modell alapú módszerek,
  - gráf alapú módszerek,

Algoritmusok osztályozása

- Ezek megközelítések tovább csoportosíthatóak:

Algoritmusok osztályozása

- Ezek megközelítések tovább csoportosíthatóak:
  - memória alalpú módszerek: hasonlóság alapú;

Algoritmusok osztályozása

- Ezek megközelítések tovább csoportosíthatóak:
  - memória alalpú módszerek: hasonlóság alapú;
  - dimenzió redukciós módszerek: mátrix faktorizácó, klaszterezés;

Algoritmusok osztályozása

- Ezek megközelítések tovább csoportosíthatóak:
  - memória alalpú módszerek: hasonlóság alapú;
  - dimenzió redukciós módszerek: mátrix faktorizácó, klaszterezés;
  - hiba minimalizációs módszerek: gépi tanulás, klaszterezés(néha), mátrix faktorizáció(néha);

Algoritmusok osztályozása

- Ezek megközelítések tovább csoportosíthatóak:
  - memória alalpú módszerek: hasonlóság alapú;
  - dimenzió redukciós módszerek: mátrix faktorizácó, klaszterezés;
  - hiba minimalizációs módszerek: gépi tanulás, klaszterezés(néha), mátrix faktorizáció(néha);
  - modell alapú megközelítések: mátrix faktorizáció, klaszterezés, gépi tanulás, valószínűségi modellek.

Algoritmusok osztályozása

- Ezek megközelítések tovább csoportosíthatóak:
  - memória alalpú módszerek: hasonlóság alapú;
  - dimenzió redukciós módszerek: mátrix faktorizácó, klaszterezés;
  - hiba minimalizációs módszerek: gépi tanulás, klaszterezés(néha), mátrix faktorizáció(néha);
  - modell alapú megközelítések: mátrix faktorizáció, klaszterezés, gépi tanulás, valószínűségi modellek.
- A feladat definíciója és az algoritmusok csoportosítása [Adomavicius-2005] alapján lett kimondva.

Memory-based user-user CF algoritmusok

- Alapötlet: $\hat{f}(u,i) = aggregate_{u' \in U_i}\left(f(u',i)\right)$, ahol U_i azoknak a felhasználóknak a halmaza, akik szavaztak az i termékre.

Memory-based user-user CF algoritmusok

- Alapötlet: $\hat{f}(u,i) = aggregate_{u' \in U_i}\left(f(u',i)\right)$, ahol U_i azoknak a felhasználóknak a halmaza, akik szavaztak az i termékre.
- Tulajdonképpen a módszer a többi felhasználó (szofisztikálabb esetekben a hasonló preferenciával rendelkező felhasználók) adott termékre tett értékeléseit ragadja meg.

Memory-based user-user CF algoritmusok

- Alapötlet: $\hat{f}(u,i) = aggregate_{u' \in U_i}\left(f(u',i)\right)$, ahol U_i azoknak a felhasználóknak a halmaza, akik szavaztak az i termékre.
- Tulajdonképpen a módszer a többi felhasználó (szofisztikálabb esetekben a hasonló preferenciával rendelkező felhasználók) adott termékre tett értékeléseit ragadja meg.
- Az egyes módszerek esetén eltér:

Memory-based user-user CF algoritmusok

- Alapötlet: $\hat{f}(u,i) = aggregate_{u' \in U_i}(f(u',i))$, ahol U_i azoknak a felhasználóknak a halmaza, akik szavaztak az i termékre.
- Tulajdonképpen a módszer a többi felhasználó (szofisztikálabb esetekben a hasonló preferenciával rendelkező felhasználók) adott termékre tett értékeléseit ragadja meg.
- Az egyes módszerek esetén eltér:
  - az aggregáció,

Memory-based user-user CF algoritmusok

- Alapötlet: $\hat{f}(u,i) = aggregate_{u' \in U_i}\left(f(u',i)\right)$, ahol U_i azoknak a felhasználóknak a halmaza, akik szavaztak az i termékre.
- Tulajdonképpen a módszer a többi felhasználó (szofisztikálabb esetekben a hasonló preferenciával rendelkező felhasználók) adott termékre tett értékeléseit ragadja meg.
- Az egyes módszerek esetén eltér:
  - az aggregáció,
  - az aggregációban használt súlyozás, amennyiben van ilyen,

Memory-based user-user CF algoritmusok

- Alapötlet: $\hat{f}(u,i) = aggregate_{u' \in U_i}(f(u',i))$, ahol U_i azoknak a felhasználóknak a halmaza, akik szavaztak az i termékre.
- Tulajdonképpen a módszer a többi felhasználó (szofisztikálabb esetekben a hasonló preferenciával rendelkező felhasználók) adott termékre tett értékeléseit ragadja meg.
- Az egyes módszerek esetén eltér:
  - az aggregáció,
  - az aggregációban használt súlyozás, amennyiben van ilyen,
  - illetve az, hogy U_i mekkora részét vesszük figyelembe.

Memory-based user-user CF algoritmusok

- Klasszikus változatok:

Memory-based user-user CF algoritmusok

- Klasszikus változatok:
  - [Resnick-1994]: $\hat{f}(u,i) = \frac{\sum_{k \in topN_{s,n}(U_i)} s_{i,k}\left(f(k,j) - avg_{q \in I_k}(f(k,q))\right)}{\sum_{k \in topN_{s,n}(U_i)} |s_{i,k}|} + avg_{q \in I_i}\left(f\left(i,q\right)\right)$,

    ahol $s_{i,k}$ az i. és k. user Pearson korrelációja, $topN_{s,n}(U)$ megadja az U halmaz elemeinek s szerinti rendezésben az első n elemből képzett részhalmazát.

Memory-based user-user CF algoritmusok

- Klasszikus változatok:
  - [Resnick-1994]: $\hat{f}(u,i) = \frac{\sum_{k \in topN_{s,n}(U_i)} s_{i,k}\left(f(k,j) - avg_{q \in I_k}(f(k,q))\right)}{\sum_{k \in topN_{s,n}(U_i)} |s_{i,k}|} + avg_{q \in I_i}\left(f\left(i,q\right)\right)$, ahol $s_{i,k}$ az i. és k. user Pearson korrelációja, $topN_{s,n}(U)$ megadja az U halmaz elemeinek s szerinti rendezésben az első n elemből képzett részhalmazát.
  - [Adomavicius-2005], [Zhou-2008]: $\hat{f}(u,i) = \frac{\sum_{k \in topN_{s,n}(U_i)} s_{i,k} f(k,j)}{\sum_{k \in topN_{s,n}(U_i)} |s_{i,k}|}$, ahol $s_{i,k}$ az i. és k. user koszinusz hasonlósága.

Központosított eljárások

[Konstan-1999] – memory-based user-user CF

- A cikkben különböző user-user alapú CF algoritmus változatokat hasonlítottak össze a GroupLens adatbázison (1173 user, 122 176 értékelés, 10% test user MovieLens kis adatbázisa). A változtatások:

[Konstan-1999] – memory-based user-user CF

- A cikkben különböző user-user alapú CF algoritmus változatokat hasonlítottak össze a GroupLens adatbázison (1173 user, 122 176 értékelés, 10% test user MovieLens kis adatbázisa). A változtatások:
  - aggregáció: korábban bemutatott kettő + z-score-average ([Resnick-1994] egy változata); ($\Rightarrow$ [Resnick-1994] nyer sokkal)

[Konstan-1999] – memory-based user-user CF

- A cikkben különböző user-user alapú CF algoritmus változatokat hasonlítottak össze a GroupLens adatbázison (1173 user, 122 176 értékelés, 10% test user MovieLens kis adatbázisa). A változtatások:
  - aggregáció: korábban bemutatott kettő + z-score-average ([Resnick-1994] egy változata); ($\Rightarrow$ [Resnick-1994] nyer sokkal)
  - hasonlóság: Pearson-, Spearman-korreláció, vektor hasonlóság, entropia, négyzetes különbség; ($\Rightarrow$ Pearson nyer kicsivel)

[Konstan-1999] – memory-based user-user CF

- A cikkben különböző user-user alapú CF algoritmus változatokat hasonlítottak össze a GroupLens adatbázison (1173 user, 122 176 értékelés, 10% test user MovieLens kis adatbázisa). A változtatások:
  - aggregáció: korábban bemutatott kettő + z-score-average ([Resnick-1994] egy változata); ($\Rightarrow$ [Resnick-1994] nyer sokkal)
  - hasonlóság: Pearson-, Spearman-korreláció, vektor hasonlóság, entropia, négyzetes különbség; ($\Rightarrow$ Pearson nyer kicsivel)
  - alacsony "együttértékelésű" hasonlóságok leértékelése: van, nincs; ($\Rightarrow$ sokat segít ha van)

[Konstan-1999] – memory-based user-user CF

- A változtatások (folyt.):

[Konstan-1999] – memory-based user-user CF

- A változtatások (folyt.):
  - szomszédség mértéke.

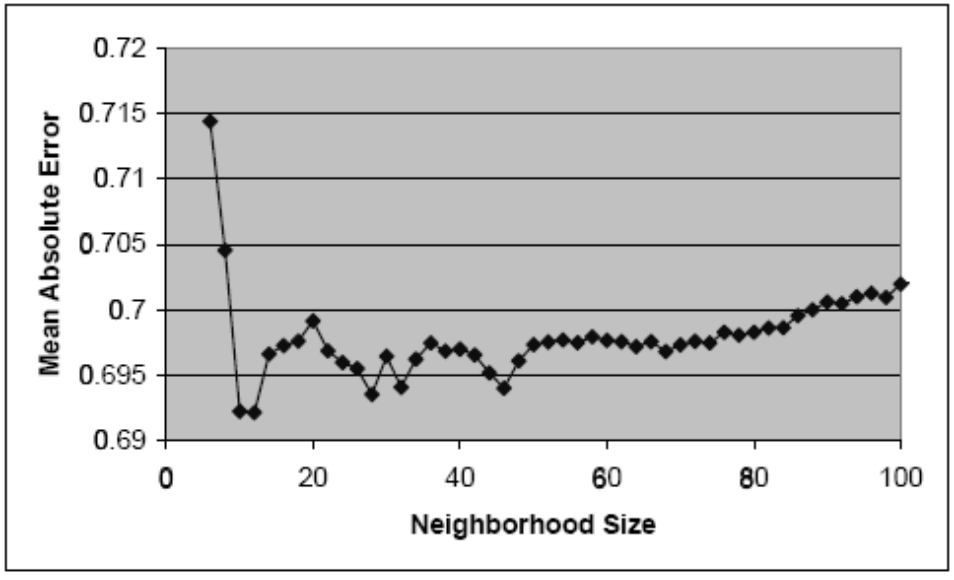


Figure 6: As you increase the size of the neighborhood, the quality of the prediction decreases, indicated by mean absolute error. This experiment was performed using Pearson correlation with $n/50$ significance weight.

Memory-based user-user CF algoritmusok

- Pozitívumok:

Memory-based user-user CF algoritmusok

- Pozitívumok:
  - Intuitív, könnyen értelmezhető és interpretálható

Memory-based user-user CF algoritmusok

- Pozitívumok:
  - Intuitív, könnyen értelmezhető és interpretálható
- Negatívumok:

Memory-based user-user CF algoritmusok

- Pozitívumok:
  - Intuitív, könnyen értelmezhető és interpretálható
- Negatívumok:
  - Érzékeny az adathalmaz (túl) ritka voltára: nagy valószínűséggel kapunk 0 hasonlóságot [Billsus-1998].

Memory-based user-user CF algoritmusok

- Pozitívumok:
  - Intuitív, könnyen értelmezhető és interpretálható
- Negatívumok:
  - Érzékeny az adathalmaz (túl) ritka voltára: nagy valószínűséggel kapunk 0 hasonlóságot [Billsus-1998].
  - Érzékeny az "new user" és "new item" problémákra [Yu-2004]

Memory-based item-item CF algoritmusok

- Alapötlet: $\hat{f}(u,i) = aggregate_{i' \in I_u}\left(f(u,i')\right)$, ahol I_u az u user által értékelt termékek halmaza.

Memory-based item-item CF algoritmusok

- Alapötlet: $\hat{f}(u,i) = aggregate_{i' \in I_u}\left(f(u,i')\right)$, ahol I_u az u user által értékelt termékek halmaza.
- Ebben a megközelítésben módszer az adott felhasználó többi (szofisztikálabb esetben a hasonló karakterisztikával rendelkező termékekre) termékekre tett értékelések alapján predikál.

Memory-based item-item CF algoritmusok

- Alapötlet: $\hat{f}(u,i) = aggregate_{i' \in I_u}\left(f(u,i')\right)$, ahol I_u az u user által értékelt termékek halmaza.
- Ebben a megközelítésben módszer az adott felhasználó többi (szofisztikálabb esetben a hasonló karakterisztikával rendelkező termékekre) termékekre tett értékelések alapján predikál.
- Tulajdonképpen a user-user alapú megközelítés transzpontáltja, mégis:

Memory-based item-item CF algoritmusok

- Alapötlet: $\hat{f}(u,i) = aggregate_{i' \in I_u}\left(f(u,i')\right)$, ahol I_u az u user által értékelt termékek halmaza.
- Ebben a megközelítésben módszer az adott felhasználó többi (szofisztikálabb esetben a hasonló karakterisztikával rendelkező termékekre) termékekre tett értékelések alapján predikál.
- Tulajdonképpen a user-user alapú megközelítés transzpontáltja, mégis:
  - jobban skálázódik mint az [Sarwar-2001],

Memory-based item-item CF algoritmusok

- Alapötlet: $\hat{f}(u,i) = aggregate_{i' \in I_u}(f(u,i'))$, ahol I_u az u user által értékelt termékek halmaza.
- Ebben a megközelítésben módszer az adott felhasználó többi (szofisztikálabb esetben a hasonló karakterisztikával rendelkező termékekre) termékekre tett értékelések alapján predikál.
- Tulajdonképpen a user-user alapú megközelítés transzpontáltja, mégis:
  - jobban skálázódik mint az [Sarwar-2001],
  - egy egyszerű ötlettel kezelhető(bb)ek a nagyon ritka adatbázisok is [Sarwar-2001].

[Sarwar-2001] – memory-based item-item CF

- Alap aggregáció: $\hat{f}(u,i) = \frac{\sum_{i' \in topN_{s,n}(I_u)} s_{i,i'} f(u,i')}{\sum_{i' \in topN_{s,n}(I_u)} |s_{i,i'}|}$.

[Sarwar-2001] – memory-based item-item CF

- Alap aggregáció: $\hat{f}(u,i) = \frac{\sum_{i' \in topN_{s,n}(I_u)} s_{i,i'} f(u,i')}{\sum_{i' \in topN_{s,n}(I_u)} |s_{i,i'}|}$.
- Regression based aggregáció: $f(u,i')$ helyett $f'(u,i') = \alpha f(u,i') + \beta + \epsilon$, ahol α és β az i' item u szerinti lineáris regressziójának paraméterei, ϵ pedig az ehhez tartozó regressziós hiba.

[Sarwar-2001] – memory-based item-item CF

- Alap aggregáció: $\hat{f}(u,i) = \frac{\sum_{i' \in topN_{s,n}(I_u)} s_{i,i'} f(u,i')}{\sum_{i' \in topN_{s,n}(I_u)} |s_{i,i'}|}$.
- Regression based aggregáció: $f(u,i')$ helyett $f'(u,i') = \alpha f(u,i') + \beta + \epsilon$, ahol α és β az i' item u szerinti lineáris regressziójának paraméterei, ϵ pedig az ehhez tartozó regressziós hiba.
- Az s alternatívái:

[Sarwar-2001] – memory-based item-item CF

- Alap aggregáció: $\hat{f}(u,i) = \frac{\sum_{i' \in topN_{s,n}(I_u)} s_{i,i'} f(u,i')}{\sum_{i' \in topN_{s,n}(I_u)} |s_{i,i'}|}$.
- Regression based aggregáció: $f(u,i')$ helyett $f'(u,i') = \alpha f(u,i') + \beta + \epsilon$, ahol α és β az i' item u szerinti lineáris regressziójának paraméterei, ϵ pedig az ehhez tartozó regressziós hiba.
- Az s alternatívái:
  - koszinusz hasonlóság: $s_{i,j} = \frac{\vec{f(.,i)} \cdot \vec{f(.,j)}}{\left\|\vec{f(.,i)}\right\| \left\|\vec{f(.,j)}\right\|}$,

[Sarwar-2001] – memory-based item-item CF

- Alap aggregáció: $\hat{f}(u,i) = \frac{\sum_{i' \in topN_{s,n}(I_u)} s_{i,i'} f(u,i')}{\sum_{i' \in topN_{s,n}(I_u)} |s_{i,i'}|}$.
- Regression based aggregáció: $f(u,i')$ helyett $f'(u,i') = \alpha f(u,i') + \beta + \epsilon$, ahol α és β az i' item u szerinti lineáris regressziójának paraméterei, ϵ pedig az ehhez tartozó regressziós hiba.
- Az s alternatívái:
  - koszinusz hasonlóság: $s_{i,j} = \frac{f(\vec{.},i) \cdot f(\vec{.},j)}{\left\| f(\vec{.},i) \right\| \left\| f(\vec{.},j) \right\|}$,
  - Pearson korreláció: $s_{i,j} = \frac{\sum_{u \in U_i \cap U_j} \left(f(u,i) - avg_{q \in U_i}(f(q,i)) \right) \left(f(u,j) - avg_{q \in U_j}(f(q,j)) \right)}{\sqrt{\sum_{u \in U_i} \left(f(u,i) - avg_{q \in U_i}(f(q,i)) \right)^2} \sqrt{\sum_{u \in U_j} \left(f(u,j) - avg_{q \in U_j}(f(q,j)) \right)^2}}$,

[Sarwar-2001] – memory-based item-item CF

- Az s alternatívái (folyt.):

[Sarwar-2001] – memory-based item-item CF

- Az s alternatívái (folyt.):
  - Módosított koszinusz hasonlóság: $s_{i,j} = \frac{\sum_{u \in U_i \cap U_j} \left(f(u,i) - avg_{q \in U_i}(f(q,i))\right)\left(f(u,j) - avg_{q \in U_j}(f(q,j))\right)}{\sqrt{\sum_{u \in U_i} (f(u,i) - avg_{q \in I_u}(f(u,q)))^2} \sqrt{\sum_{u \in U_j} (f(u,j) - avg_{q \in I_u}(f(u,q)))^2}}$.

[Sarwar-2001] – memory-based item-item CF

- Az s alternatívái (folyt.):
  - Módosított koszinusz hasonlóság: $s_{i,j} = \frac{\sum_{u \in U_i \cap U_j} \left(f(u,i) - avg_{q \in U_i}(f(q,i))\right)\left(f(u,j) - avg_{q \in U_j}(f(q,j))\right)}{\sqrt{\sum_{u \in U_i} (f(u,i) - avg_{q \in I_u}(f(u,q)))^2} \sqrt{\sum_{u \in U_j} (f(u,j) - avg_{q \in I_u}(f(u,q)))^2}}$.
- A cikkben tárgyalt 6 módszert (2 predikciós eljárás, 3 hasonlósági mértékkel) MovieLens adatbázis legkisebb változatán értékelték ki.

[Sarwar-2001] – memory-based item-item CF

- Hasonlósági mértékek hatása (alap aggregáció esetén):

[Sarwar-2001] – memory-based item-item CF

- Hasonlósági mértékek hatása (alap aggregáció esetén):

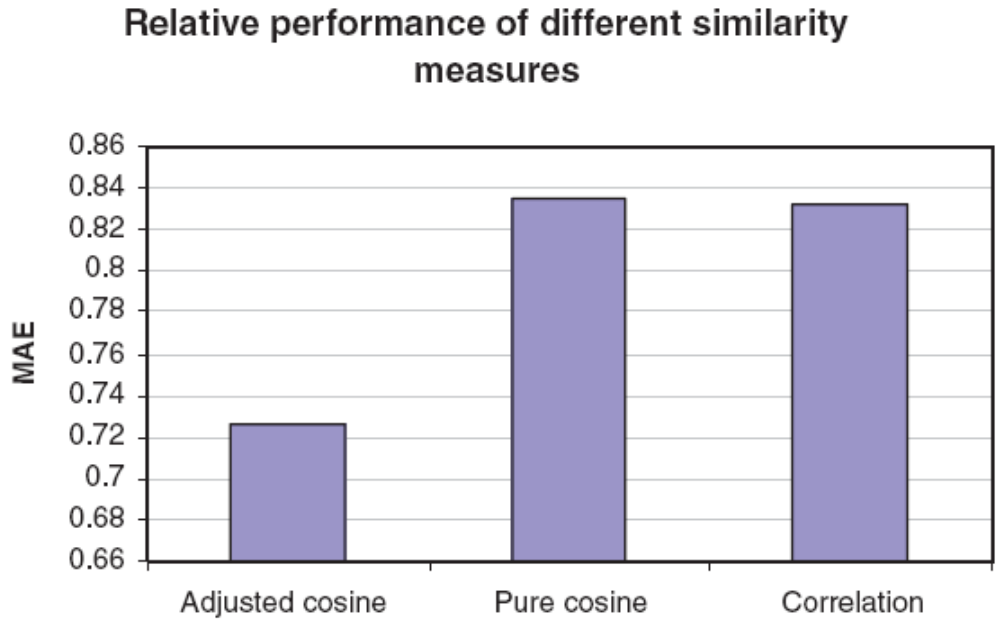


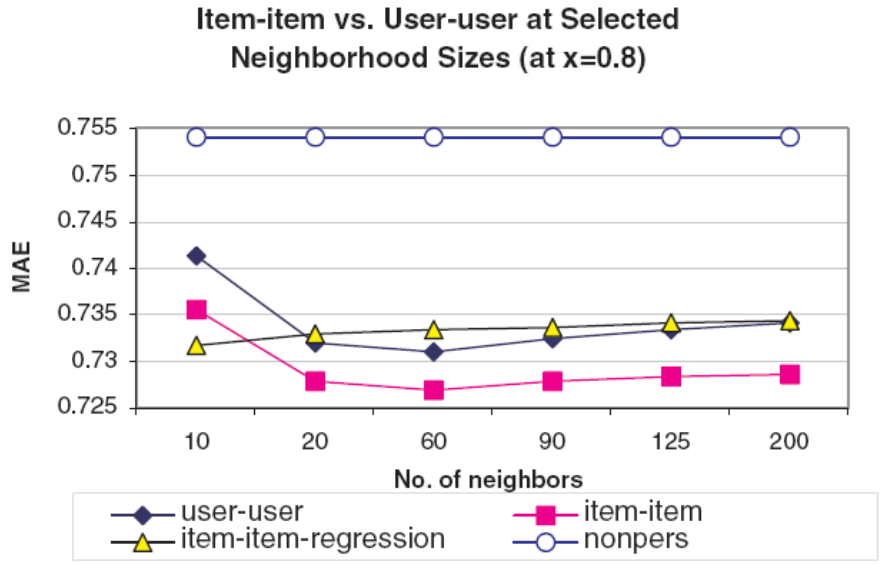
Figure 4: Impact of the similarity computation measure on item-based collaborative filtering algorithm.

[Sarwar-2001] – memory-based item-item CF

- Szomszédság méretének és a train/test hányados hatásai:

[Sarwar-2001] – memory-based item-item CF

- Szomszédság méretének és a train/test hányados hatásai:



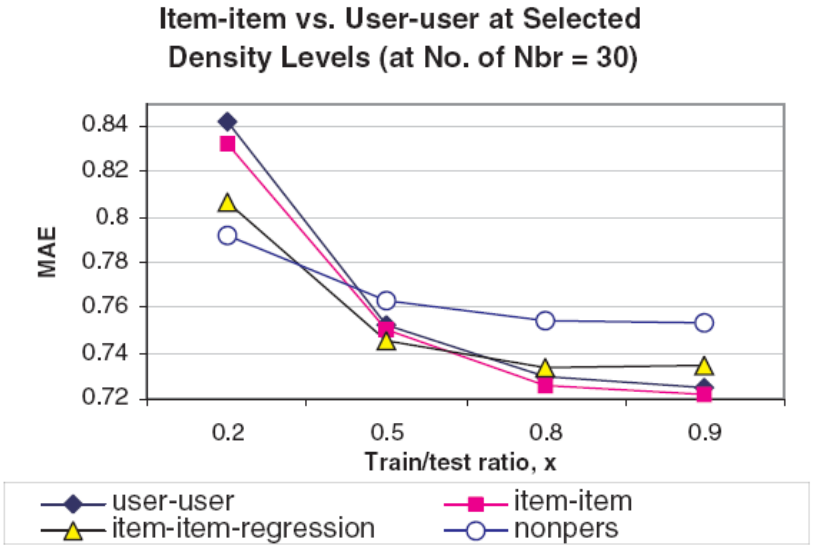


Figure 6: Comparison of prediction quality of *item-item* and *user-user* collaborative filtering algorithms. We compare prediction qualities at $x = 0.2, 0.5, 0.8$ and 0.9.

Memory-based item-item CF algoritmusok

- Pozitívumok:

Memory-based item-item CF algoritmusok

- Pozitívumok:
  - Szintén intuitív, könnyen értelmezhető és interpretálható

Memory-based item-item CF algoritmusok

- Pozitívumok:
  - Szintén intuitív, könnyen értelmezhető és interpretálható
  - Empirikus elemzésekkel kimutatták, hogy szignifikánsan jobb, mint a hasonló user-user alapú megközelítések [Sarwar-2001].

Memory-based item-item CF algoritmusok

- Pozitívumok:
  - Szintén intuitív, könnyen értelmezhető és interpretálható
  - Empirikus elemzésekkel kimutatták, hogy szignifikánsan jobb, mint a hasonló user-user alapú megközelítések [Sarwar-2001].
- Negatívumok:

Memory-based item-item CF algoritmusok

- Pozitívumok:
  - Szintén intuitív, könnyen értelmezhető és interpretálható
  - Empirikus elemzésekkel kimutatták, hogy szignifikánsan jobb, mint a hasonló user-user alapú megközelítések [Sarwar-2001].
- Negatívumok:
  - Érzékeny az adathalmaz (túl) ritka voltára, habár különböző technikákkal (smoothing) javítottak ezen [Hu-2006].

Memory-based item-item CF algoritmusok

- Pozitívumok:
  - Szintén intuitív, könnyen értelmezhető és interpretálható
  - Empirikus elemzésekkel kimutatták, hogy szignifikánsan jobb, mint a hasonló user-user alapú megközelítések [Sarwar-2001].
- Negatívumok:
  - Érzékeny az adathalmaz (túl) ritka voltára, habár különböző technikákkal (smoothing) javítottak ezen [Hu-2006].
  - Érzékeny az "new user" és "new item" problémákra, azonban ezen is segítenek (valamennyit) a smoothing technikák [Yu-2004], [Hu-2006].

Gráf alapú megközelítések

- Alapötlet: a rendelkezésre álló $f(u, i)$ adatokból képezzünk gráfot és a gráfelmélet technikáit segítségül hívva.

Gráf alapú megközelítések

- Alapötlet: a rendelkezésre álló $f(u, i)$ adatokból képezzünk gráfot és a gráfelmélet technikáit segítségül hívva.
- Megközelítések:

Gráf alapú megközelítések

- Alapötlet: a rendelkezésre álló $f(u, i)$ adatokból képezzünk gráfot és a gráfelmélet technikáit segítségül hívva.
- Megközelítések:
  - Páros gráf: az csomópontok halmaza $U \cup I$, élek csak $u \in U$ és $i \in I$ csomópontok között mennek (páros gráf), pontosan azon u és i csomópontok között megy él, ahol $f(u, i)$ ismert [Huang-2004], [Zhou-2008].

Gráf alapú megközelítések

- Alapötlet: a rendelkezésre álló $f(u, i)$ adatokból képezzünk gráfot és a gráfelmélet technikáit segítségül hívva.
- Megközelítések:
  - Páros gráf: az csomópontok halmaza $U \cup I$, élek csak $u \in U$ és $i \in I$ csomópontok között mennek (páros gráf), pontosan azon u és i csomópontok között megy él, ahol $f(u, i)$ ismert [Huang-2004], [Zhou-2008].
  - User-user gráf: a csomópontok halmaza U, élek hasonlóság vagy egyéb heurisztika alapján képződnek (amik az hasonló preferenciát, vagy az egymás alapján történő predikálhatóságot modellezik) [Aggarwal-1999].

[Aggarwal-1999] – Horting hatches an egg...

- A cikkben használt megközelítés irányított user-user gráfot épít, amely két relációt ír le:

[Aggarwal-1999] – Horting hatches an egg...

- A cikkben használt megközelítés irányított user-user gráfot épít, amely két relációt ír le:
  - $hortes(u_1, u_2) =$

$$\begin{cases} \dfrac{|I_{u_1} \cap I_{u_2}|}{|I_{u_1}|} \geq F, \text{ ahol } F \text{ előre adott küszöb, VAGY} \\ |I_{u_1} \cap I_{u_2}| \geq G, \text{ ahol } g \text{ előre adott küszöb;} \end{cases}$$

[Aggarwal-1999] – Horting hatches an egg...

- A cikkben használt megközelítés irányított user-user gráfot épít, amely két relációt ír le:
  - $hortes(u_1, u_2) =$

$$\begin{cases} \frac{|I_{u_1} \cap I_{u_2}|}{|I_{u_1}|} \geq F, \text{ ahol } F \text{ előre adott küszöb, VAGY} \\ |I_{u_1} \cap I_{u_2}| \geq G, \text{ ahol } g \text{ előre adott küszöb;} \end{cases}$$

  - $predictable(u_1, u2) =$

$$\begin{cases} hortes(u2, u1) \text{ ÉS} \\ \exists T_{s,t} : \frac{\sum_{j \in I_{u_1} \cap I_{u_2}} |f(u_1, j) - T_{s,t}(f(u_2, j))|}{|I_{u_1} \cap I_{u_2}|} < H, \ H \text{ konst.} \end{cases}$$

[Aggarwal-1999] – Horting hatches an egg...

- A $predictable$ relációban szereplő $T_{s,t}(f(u_2, j)) = sf(u_2, j) + t$ lineáris leképezés a bemenet értékeléshez tartozó felhasználó "viselkedési modelljét" határozza meg, egy másik (u_1) felhasználóhoz képest, azaz:

[Aggarwal-1999] – Horting hatches an egg...

- A $predictable$ relációban szereplő $T_{s,t}(f(u_2, j)) = sf(u_2, j) + t$ lineáris leképezés a bemenet értékeléshez tartozó felhasználó "viselkedési modelljét" határozza meg, egy másik (u_1) felhasználóhoz képest, azaz:
  - $s = 1$ és $t = 0$ esetén az u_1 és u_2 felhasználók viselekedése *azonos*,

[Aggarwal-1999] – Horting hatches an egg...

- A $predictable$ relációban szereplő $T_{s,t}(f(u_2, j)) = sf(u_2, j) + t$ lineáris leképezés a bemenet értékeléshez tartozó felhasználó "viselkedési modelljét" határozza meg, egy másik (u_1) felhasználóhoz képest, azaz:
  - $s = 1$ és $t = 0$ esetén az u_1 és u_2 felhasználók viselekedése *azonos*,
  - $s = -1$ és $t = 0$ esetén az u_1 és u_2 felhasználók viselekedése *ellentétes*...

[Aggarwal-1999] – Horting hatches an egg...

- A $predictable$ relációban szereplő $T_{s,t}(f(u_2, j)) = sf(u_2, j) + t$ lineáris leképezés a bemenet értékeléshez tartozó felhasználó "viselkedési modelljét" határozza meg, egy másik (u_1) felhasználóhoz képest, azaz:
  - $s = 1$ és $t = 0$ esetén az u_1 és u_2 felhasználók viselekedése *azonos*,
  - $s = -1$ és $t = 0$ esetén az u_1 és u_2 felhasználók viselekedése *ellentétes*...
- Az $s \in \{+1, -1\}$ és $t \in \{2, 2v, 1 - v, v - 1\}$ értékek egy szűk halmazból kerülnek kiválasztásra, így a megfelelő $T_{s,t}$ meghatározása kimerítő kereséssel is hatékony.

[Aggarwal-1999] – Horting hatches an egg...

- A $predictable$ relációban szereplő $T_{s,t}(f(u_2, j)) = sf(u_2, j) + t$ lineáris leképezés a bemenet értékeléshez tartozó felhasználó "viselkedési modelljét" határozza meg, egy másik (u_1) felhasználóhoz képest, azaz:
  - $s = 1$ és $t = 0$ esetén az u_1 és u_2 felhasználók viselekedése *azonos*,
  - $s = -1$ és $t = 0$ esetén az u_1 és u_2 felhasználók viselekedése *ellentétes*...
- Az $s \in \{+1, -1\}$ és $t \in \{2, 2v, 1 - v, v - 1\}$ értékek egy szűk halmazból kerülnek kiválasztásra, így a megfelelő $T_{s,t}$ meghatározása kimerítő kereséssel is hatékony.
- A $predictalble$ reláció alapján user-user gráf megépítése.

[Aggarwal-1999] – Horting hatches an egg...

- $\hat{f}\,(u, j)$ meghatározása:

[Aggarwal-1999] – Horting hatches an egg...

- $\hat{f}(u, j)$ meghatározása:
  - legrövidebb $u \rightarrow u_1 \rightarrow \ldots \rightarrow u_l$, úgy hogy $f(u_l, j)$ létezzen,

[Aggarwal-1999] – Horting hatches an egg...

- $\hat{f}(u, j)$ meghatározása:
  - legrövidebb $u \rightarrow u_1 \rightarrow \ldots \rightarrow u_l$, úgy hogy $f(u_l, j)$ létezzen,
  - $\hat{f}(u, j) = T_{s_1^*, t_1^*} \circ \ldots \circ T_{s_l^*, t_l^*}(f(u_l, j))$.

[Aggarwal-1999] – Horting hatches an egg...

- $\hat{f}(u, j)$ meghatározása:
  - legrövidebb $u \rightarrow u_1 \rightarrow \ldots \rightarrow u_l$, úgy hogy $f(u_l, j)$ létezzen,
  - $\hat{f}(u, j) = T_{s_1^*, t_1^*} \circ \ldots \circ T_{s_l^*, t_l^*}(f(u_l, j))$.
- Kiértékelés base-line nélkül 3 különböző (számomra ismeretlen) adtbázison...

[Huang-2004] – páros gráf alapú megközelítés

- A cikkben kidolgozott módszert Kleinberg HITS algoritmusa inspirálta.

[Huang-2004] – páros gráf alapú megközelítés

- A cikkben kidolgozott módszert Kleinberg HITS algoritmusa inspirálta.
- Adott u_0 felhasználó esetén definiáljuk a következő score-okat:

[Huang-2004] – páros gráf alapú megközelítés

- A cikkben kidolgozott módszert Kleinberg HITS algoritmusa inspirálta.
- Adott u_0 felhasználó esetén definiáljuk a következő score-okat:
  - $\forall i \in I : pr\left(u_0, i\right)$ - termék reprezentativitás: megmondja, hogy u_0 felhasználót mennyire érdekli az i termék,

[Huang-2004] – páros gráf alapú megközelítés

- A cikkben kidolgozott módszert Kleinberg HITS algoritmusa inspirálta.
- Adott u_0 felhasználó esetén definiáljuk a következő score-okat:
  - $\forall i \in I : pr\,(u_0, i)$ - termék reprezentativitás: megmondja, hogy u_0 felhasználót mennyire érdekli az i termék,
  - $\forall u \in U : cr\,(u_0, u)$ - felhasználó reprezentativitás: megmondja, hogy u_0 felhasználóhoz mennyire hasonlít az u user.

[Huang-2004] – páros gráf alapú megközelítés

- A reprezentativitások segítségével felírhatóak a következő mágrixok:

[Huang-2004] – páros gráf alapú megközelítés

- A reprezentativitások segítségével felírhatóak a következő mágrixok:
  - $PR = \{pr(u,i)\}_{i,j} \in \mathrm{R}^{|U|\times|I|}$

[Huang-2004] – páros gráf alapú megközelítés

- A reprezentativitások segítségével felírhatóak a következő mágrixok:
  - $PR = \{pr(u,i)\}_{i,j} \in \mathrm{R}^{|U|\times|I|}$
  - $CR = \{cr(u_1,u_2)\}_{i,j} \in \mathrm{R}^{|U|\times|U|}$

[Huang-2004] – páros gráf alapú megközelítés

- A reprezentativitások segítségével felírhatóak a következő mágrixok:
  - $PR = \{pr(u,i)\}_{i,j} \in \mathrm{R}^{|U|\times|I|}$
  - $CR = \{cr(u_1,u_2)\}_{i,j} \in \mathrm{R}^{|U|\times|U|}$
- A reprezentativitások definíciója rekurziv (Kleinber szerű értelemben), azaz:

[Huang-2004] – páros gráf alapú megközelítés

- A reprezentativitások segítségével felírhatóak a következő mágrixok:
  - $PR = \{pr(u,i)\}_{i,j} \in \mathrm{R}^{|U|\times|I|}$
  - $CR = \{cr(u_1,u_2)\}_{i,j} \in \mathrm{R}^{|U|\times|U|}$
- A reprezentativitások definíciója rekurziv (Kleinber szerű értelemben), azaz:
  - $PR = h\,(CR)$

[Huang-2004] – páros gráf alapú megközelítés

- A reprezentativitások segítségével felírhatóak a következő mágrixok:
  - $PR = \{pr(u,i)\}_{i,j} \in \mathrm{R}^{|U|\times|I|}$
  - $CR = \{cr(u_1,u_2)\}_{i,j} \in \mathrm{R}^{|U|\times|U|}$
- A reprezentativitások definíciója rekurziv (Kleinber szerű értelemben), azaz:
  - $PR = h\left(CR\right)$
  - $CR = g\left(PR\right)$

[Huang-2004] – páros gráf alapú megközelítés

- A reprezentativitások segítségével felírhatóak a következő mágrixok:
  - $PR = \{pr(u,i)\}_{i,j} \in \mathrm{R}^{|U|\times|I|}$
  - $CR = \{cr(u_1,u_2)\}_{i,j} \in \mathrm{R}^{|U|\times|U|}$
- A reprezentativitások definíciója rekurziv (Kleinber szerű értelemben), azaz:
  - $PR = h\left(CR\right)$
  - $CR = g\left(PR\right)$
- Amennyiben $h(CR) = CR^T A^T$ és $g(PR) = A^T PR$, akkor a HITS algoritmust kapjuk vissza a PR és CR megfelelő sor- illetve oszlopvektoraira.

[Huang-2004] – páros gráf alapú megközelítés

- Amennyiben:

[Huang-2004] – páros gráf alapú megközelítés

- Amennyiben:
  - $h(CR)_{i,k} = \frac{\sum_{j \text{ és } a_{j,k} \neq null} a_{j,k} cr(i,j)}{\sum_{j \text{ és } a_{j,k} \neq null} a_{j,k}}$ és

[Huang-2004] – páros gráf alapú megközelítés

- Amennyiben:
  - $h(CR)_{i,k} = \frac{\sum_{j \text{ és } a_{j,k} \neq null} a_{j,k} cr(i,j)}{\sum_{j \text{ és } a_{j,k} \neq null} a_{j,k}}$ és
  - $g(PR)_{i,t} = \frac{\sum_{j \text{ és } a_{j,k} \neq null} a_{t,j} pr(i,j)}{\sum_{j \text{ és } a_{t,j} \neq null} a_{t,j}}$,

[Huang-2004] – páros gráf alapú megközelítés

- Amennyiben:
  - $h(CR)_{i,k} = \frac{\sum_{j \text{ és } a_{j,k} \neq null} a_{j,k} cr(i,j)}{\sum_{j \text{ és } a_{j,k} \neq null} a_{j,k}}$ és
  - $g(PR)_{i,t} = \frac{\sum_{j \text{ és } a_{j,k} \neq null} a_{t,j} pr(i,j)}{\sum_{j \text{ és } a_{t,j} \neq null} a_{t,j}}$,

  akkor a PageRank-ben implementált random walk modellt kapjuk.

[Huang-2004] – páros gráf alapú megközelítés

- Eredmények: F-measure-ben, Chinese online book database, 9695 item, 2000 user, 18 771 link.

[Huang-2004] – páros gráf alapú megközelítés

- Eredmények: F-measure-ben, Chinese online book database, 9695 item, 2000 user, 18 771 link.

Algorithm	Training Set Percentage			
	10%	20%	40%	70%
User-based	0.001905	0.00365	0.009078	0.013759
Item-based	0.005249	0.001622	0.000001	0.004348
Link Analysis	**0.010042**	**0.028494**	**0.023346**	**0.045226**

Machine Learning Model based methods

- Alapötlet: a rendelkezésre álló $f(u, i)$ adathalmazból, a felhasználó profilokból valamint a termék karakterisztikákból készítsünk jellemzővektorokat, az predikciót osztálycímkének tekintve tanítsuk felügyelt gépi tanuló modellt.

Machine Learning Model based methods

- Alapötlet: a rendelkezésre álló $f(u, i)$ adathalmazból, a felhasználó profilokból valamint a termék karakterisztikákból készítsünk jellemzővektorokat, az predikciót osztálycímkének tekintve tanítsuk felügyelt gépi tanuló modellt.
- Jellemző vektorokat lehet,

Machine Learning Model based methods

- Alapötlet: a rendelkezésre álló $f(u, i)$ adathalmazból, a felhasználó profilokból valamint a termék karakterisztikákból készítsünk jellemzővektorokat, az predikciót osztálycímkének tekintve tanítsuk felügyelt gépi tanuló modellt.
- Jellemző vektorokat lehet,
  - minden felhasznlóra és item-re külön kinyerni,

Machine Learning Model based methods

- Alapötlet: a rendelkezésre álló $f(u, i)$ adathalmazból, a felhasználó profilokból valamint a termék karakterisztikákból készítsünk jellemzővektorokat, az predikciót osztálycímkének tekintve tanítsuk felügyelt gépi tanuló modellt.
- Jellemző vektorokat lehet,
  - minden felhasznlóra és item-re külön kinyerni,
  - felhasználonként itemek egy csoportjához (klaszteréhez),

Machine Learning Model based methods

- Alapötlet: a rendelkezésre álló $f(u, i)$ adathalmazból, a felhasználó profilokból valamint a termék karakterisztikákból készítsünk jellemzővektorokat, az predikciót osztálycímkének tekintve tanítsuk felügyelt gépi tanuló modellt.
- Jellemző vektorokat lehet,
  - minden felhasznlóra és item-re külön kinyerni,
  - felhasználonként itemek egy csoportjához (klaszteréhez),
  - ezekből képzett hibrid módszerrel.

[Park-2008] – ML model

- A cikkben azt vizsgálták, hogy a különböző gépi tanulási modellekre hogyan hat az item értékelési gyakoriság,

[Park-2008] – ML model

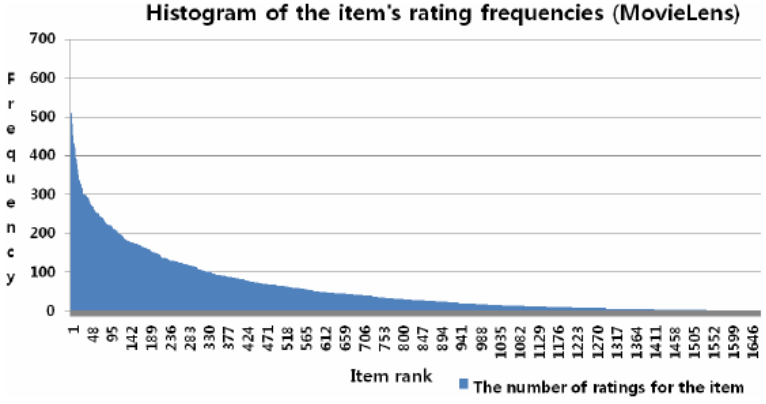
- A cikkben azt vizsgálták, hogy a különböző gépi tanulási modellekre hogyan hat az item értékelési gyakoriság,
- illetve arra kerestek választ, hogy a gépi tanulási modelek három felhasználása közül melyik a legjobb.

[Park-2008] – ML model

- A cikkben azt vizsgálták, hogy a különböző gépi tanulási modellekre hogyan hat az item értékelési gyakoriság,
- illetve arra kerestek választ, hogy a gépi tanulási modelek három felhasználása közül melyik a legjobb.
- Észrevétel: az item értékelési gyakoriság long-tail eloszlású

[Park-2008] – ML model

- A cikkben azt vizsgálták, hogy a különböző gépi tanulási modellekre hogyan hat az item értékelési gyakoriság,
- illetve arra kerestek választ, hogy a gépi tanulási modelek három felhasználása közül melyik a legjobb.
- Észrevétel: az item értékelési gyakoriság long-tail eloszlású

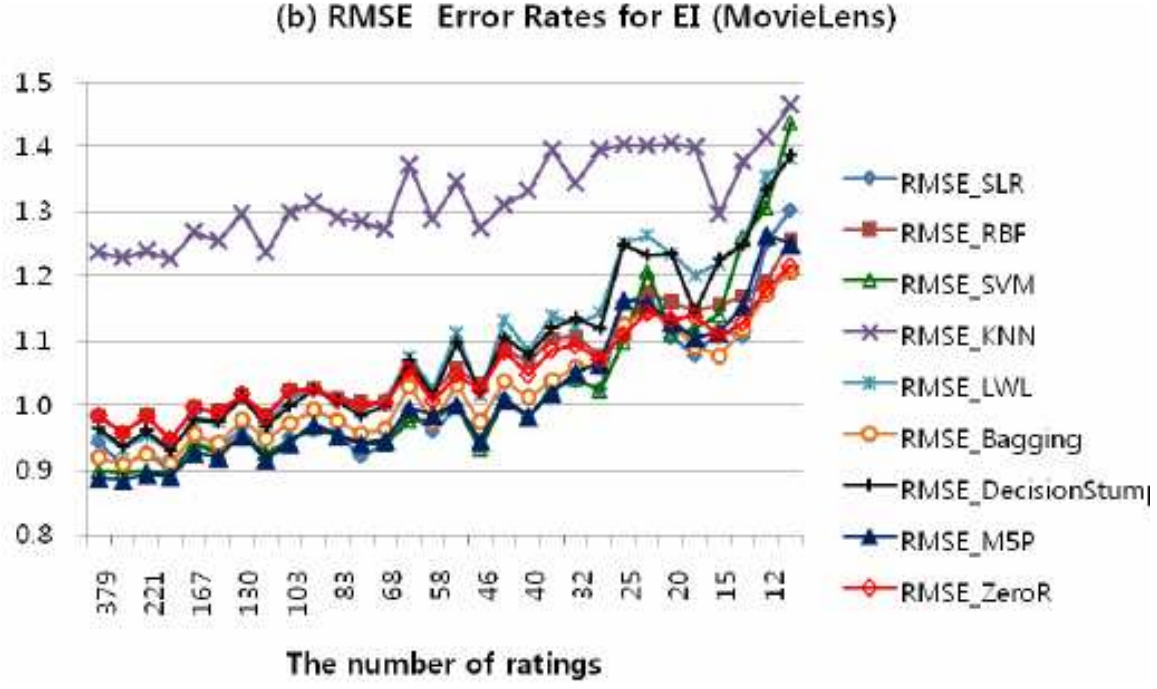


[Park-2008] – ML model

- Következmény: Ha minden egyes item-re külön készítünk modellt, akkor a modellek hatékonysága az item értékelések számának csökkenésével arányosan romlik.

[Park-2008] – ML model

- Következmény: Ha minden egyes item-re külön készítünk modellt, akkor a modellek hatékonysága az item értékelések számának csökkenésével arányosan romlik.



[Park-2008] – ML model

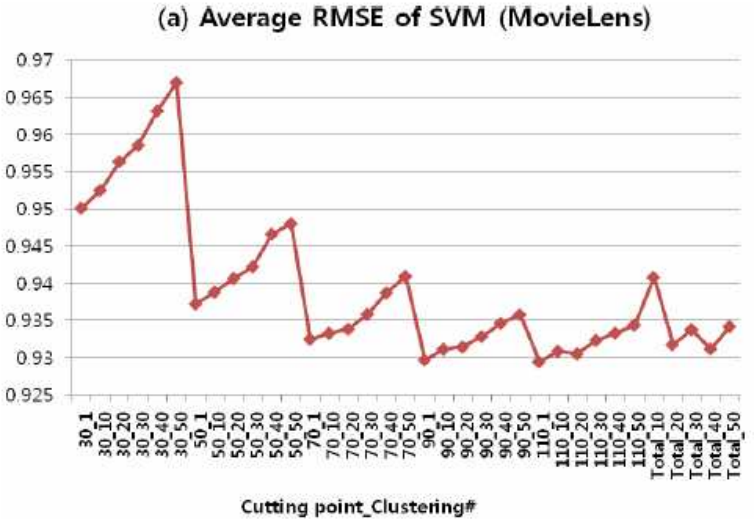
- Ha az itemeket klaszterezzük az segít, de így a gyakran értékelt itemek is klasztereződnek így veszítnek informáltságukból.

[Park-2008] – ML model

- Ha az itemeket klaszterezzük az segít, de így a gyakran értékelt itemek is klasztereződnek így veszítnek informáltságukból.
- Megoldás hibrid mószer, head-ben minden item-hez külön modell, tail-ben klaszterezés és klaszterenkénti modell.

[Park-2008] – ML model

- Ha az itemeket klaszterezzük az segít, de így a gyakran értékelt itemek is klasztereződnek így veszítnek informáltságukból.
- Megoldás hibrid mószer, head-ben minden item-hez külön modell, tail-ben klaszterezés és klaszterenkénti modell.



Dimensionality reduction based approches

- Alapötlet: a rendelkezésünkre álló ajánlásokból képzett mátrix dimenzióját csökkentsük, ezzel tömörítve azt hatékonyabban tudunk távolságot számolni, a számolt távolságértékek megbízhatóbbak lesznek.

Dimensionality reduction based approches

- Alapötlet: a rendelkezésünkre álló ajánlásokból képzett mátrix dimenzióját csökkentsük, ezzel tömörítve azt hatékonyabban tudunk távolságot számolni, a számolt távolságértékek megbízhatóbbak lesznek.
- Leggyakrabban használt eljárások:

Dimensionality reduction based approches

- Alapötlet: a rendelkezésünkre álló ajánlásokból képzett mátrix dimenzióját csökkentsük, ezzel tömörítve azt hatékonyabban tudunk távolságot számolni, a számolt távolságértékek megbízhatóbbak lesznek.
- Leggyakrabban használt eljárások:
  - LSI [Billsus-1998],

Dimensionality reduction based approches

- Alapötlet: a rendelkezésünkre álló ajánlásokból képzett mátrix dimenzióját csökkentsük, ezzel tömörítve azt hatékonyabban tudunk távolságot számolni, a számolt távolságértékek megbízhatóbbak lesznek.
- Leggyakrabban használt eljárások:
  - LSI [Billsus-1998],
  - PCA [Goldberg-2001].

Dimensionality reduction based approches

- Alapötlet: a rendelkezésünkre álló ajánlásokból képzett mátrix dimenzióját csökkentsük, ezzel tömörítve azt hatékonyabban tudunk távolságot számolni, a számolt távolságértékek megbízhatóbbak lesznek.
- Leggyakrabban használt eljárások:
  - LSI [Billsus-1998],
  - PCA [Goldberg-2001].
- Vélemények megoszlanak a hasznukról [Sarwar-2000]:

Dimensionality reduction based approches

- Alapötlet: a rendelkezésünkre álló ajánlásokból képzett mátrix dimenzióját csökkentsük, ezzel tömörítve azt hatékonyabban tudunk távolságot számolni, a számolt távolságértékek megbízhatóbbak lesznek.
- Leggyakrabban használt eljárások:
  - LSI [Billsus-1998],
  - PCA [Goldberg-2001].
- Vélemények megoszlanak a hasznukról [Sarwar-2000]:

> recommendations. More advanced techniques can be applied to achieve dimensionality reduction. Examples are statistical techniques such as Principle Component Analysis (PCA) [8] and information retrieval techniques such as Latent Semantic Indexing (LSI) [3, 22]. Empirical studies indicate that dimensionality reduction can improve recommendation quality significantly in some applications, but performs poorly in others [22]. The dimensionality reduction approach addresses the

Néhány összehasonlító eredmény

Table 2. Experimental results for six collaborative-filtering algorithms.*

Measure	Algorithm	Data set						Average algorithm rank
		Apparel		Book		Movie		
		Reduced	Unreduced	Reduced	Unreduced	Reduced	Unreduced	
F-measure	User-based	0.0072	0.0153	0.0165	0.0320	**0.0458**	0.0634	4.33
	Item-based	0.0077	0.0166	0.0072	0.0142	**0.0473**	**0.0769**	4.17
	Dimensionality-reduction	0.0088	0.0109	0.0157	0.0305	0.0386	0.0528	5.33
	Generative-model	**0.0113**	0.0107	**0.0454**	**0.0406**	0.0362	0.0447	4.00
	Spreading-activation	**0.0111**	**0.0219**	0.0320	0.0362	0.0426	0.0583	3.67
	Link-analysis	**0.0144**	**0.0224**	0.0349	**0.0415**	**0.0466**	0.0605	**1.83**
	Top-*N* most popular	**0.0113**	0.0100	**0.0431**	**0.0405**	0.0357	0.0425	4.67

Elosztott módszerek

[Tveit-2001] Szavazati lista terjesztés

- A peer-ek egymásnak küldött üzeneteik a szavazatvektorok

[Tveit-2001] Szavazati lista terjesztés

- A peer-ek egymásnak küldött üzeneteik a szavazatvektorok
- A hatékonyság növelése érdekében

[Tveit-2001] Szavazati lista terjesztés

- A peer-ek egymásnak küldött üzeneteik a szavazatvektorok
- A hatékonyság növelése érdekében
  - A vektorok ritka reprezentációban kerülnek átküldésre

[Tveit-2001] Szavazati lista terjesztés

- A peer-ek egymásnak küldött üzeneteik a szavazatvektorok
- A hatékonyság növelése érdekében
  - A vektorok ritka reprezentációban kerülnek átküldésre
  - A ritka reprezentációt tömörítik (binary interpolative compression [Moffat-1996])

[Tveit-2001] Szavazati lista terjesztés

- A peer-ek egymásnak küldött üzeneteik a szavazatvektorok
- A hatékonyság növelése érdekében
  - A vektorok ritka reprezentációban kerülnek átküldésre
  - A ritka reprezentációt tömörítik (binary interpolative compression [Moffat-1996])
- Ezen vektor alapján hasonlóság számolása (Pearson correlation vagy Cosine measure)

[Tveit-2001] Szavazati lista terjesztés

- A peer-ek egymásnak küldött üzeneteik a szavazatvektorok
- A hatékonyság növelése érdekében
  - A vektorok ritka reprezentációban kerülnek átküldésre
  - A ritka reprezentációt tömörítik (binary interpolative compression [Moffat-1996])
- Ezen vektor alapján hasonlóság számolása (Pearson correlation vagy Cosine measure)
  - ha ez meghalad egy τ korlátot, akkor visszaküldi a saját vektorát a feladónak

[Tveit-2001] Szavazati lista terjesztés

- A peer-ek egymásnak küldött üzeneteik a szavazatvektorok
- A hatékonyság növelése érdekében
  - A vektorok ritka reprezentációban kerülnek átküldésre
  - A ritka reprezentációt tömörítik (binary interpolative compression [Moffat-1996])
- Ezen vektor alapján hasonlóság számolása (Pearson correlation vagy Cosine measure)
  - ha ez meghalad egy τ korlátot, akkor visszaküldi a saját vektorát a feladónak
  - ha nem, akkor a kapott vektort továbbítja a szomszédainak.

[Tveit-2001] Szavazati lista terjesztés (2)

- Pl.: $[10:0, 22:2, 37:8]$ és $[10:0, 22:2, 24:9, 37:8]$, akkor az ajánlás a $24:9$ súlyozva

[Tveit-2001] Szavazati lista terjesztés (2)

- Pl.: $[10:0, 22:2, 37:8]$ és $[10:0, 22:2, 24:9, 37:8]$, akkor az ajánlás a $24:9$ súlyozva
- Ha elég sok vektor érkezett vissza, akkor a peer rangsorolja azokat és megteszi az ajánlást

[Tveit-2001] Szavazati lista terjesztés (2)

- Pl.: $[10:0, 22:2, 37:8]$ és $[10:0, 22:2, 24:9, 37:8]$, akkor az ajánlás a $24:9$ súlyozva
- Ha elég sok vektor érkezett vissza, akkor a peer rangsorolja azokat és megteszi az ajánlást
- A hálózat dinamikája: ha egy peer két szomszédjától nagyon hasonló vektorokat kap, akkor megkérdezi őket, hogy akarnak-e közvetlen kapcsolatot egymás között

[Bickson-2007] Ráták szétterjesztése

- Az általuk definált veszteség fgv.:
  $min_x(\sum_{i \in V}(x_i - y_i)^2 + \beta \sum_{w_{ij} \in E} w_{ij}(x_i - x_j)^2)$

[Bickson-2007] Ráták szétterjesztése

- Az általuk definált veszteség fgv.:
  $min_x(\sum_{i \in V}(x_i - y_i)^2 + \beta \sum_{w_{ij} \in E} w_{ij}(x_i - x_j)^2)$
- Minden egyes peer-hez megadták a többi peer fontossági sorrendjét a Gaussian Belief Propagation algoritmus [Johnson-2006] peer-enkénti párhuzamos futtatásával a szociális háló topológiája alapján

[Bickson-2007] Ráták szétterjesztése

- Az általuk definált veszteség fgv.:
  $min_x(\sum_{i \in V}(x_i - y_i)^2 + \beta \sum_{w_{ij} \in E} w_{ij}(x_i - x_j)^2)$
- Minden egyes peer-hez megadták a többi peer fontossági sorrendjét a Gaussian Belief Propagation algoritmus [Johnson-2006] peer-enkénti párhuzamos futtatásával a szociális háló topológiája alapján
- Az ajánlást pedig a Consensus Propagation algoritmus [Moallemi-2006] segítségével számolták, amely a hálózat átlagát számolta, hogy minimalizálja a fent definiált hibát

[Bickson-2007] Ráták szétterjesztése (2)

- Ezen algoritmust (CPA) módosították, hogy

[Bickson-2007] Ráták szétterjesztése (2)

- Ezen algoritmust (CPA) módosították, hogy
  - a β paraméterrel hangolható legyen, hogy a peerek mennyire legyenek érzékenyek szomszédokra

[Bickson-2007] Ráták szétterjesztése (2)

- Ezen algoritmust (CPA) módosították, hogy
  - a β paraméterrel hangolható legyen, hogy a peerek mennyire legyenek érzékenyek szomszédokra
  - az algoritmus kezelje a hiányzó értékeket (még szavazat nélküli itemek)

[Bickson-2007] Ráták szétterjesztése (2)

- Ezen algoritmust (CPA) módosították, hogy
  - a β paraméterrel hangolható legyen, hogy a peerek mennyire legyenek érzékenyek szomszédokra
  - az algoritmus kezelje a hiányzó értékeket (még szavazat nélküli itemek)
  - valamint ne kelljen az élsúlyoknak egyenletes eloszlást követniük

[Bickson-2007] Ráták szétterjesztése (2)

- Ezen algoritmust (CPA) módosították, hogy
  - a β paraméterrel hangolható legyen, hogy a peerek mennyire legyenek érzékenyek szomszédokra
  - az algoritmus kezelje a hiányzó értékeket (még szavazat nélküli itemek)
  - valamint ne kelljen az élsúlyoknak egyenletes eloszlást követniük
- Az algoritmus kiterjeszthető a személyre szabott PageRank és az Information Centrality módszerekre a megfelelő input paraméterekkel

[Georgios-2006] Kérés szétterjesztése

- Bizalmi overlay háló építése a [Pitsilis-2004] cikk alapján

[Georgios-2006] Kérés szétterjesztése

- Bizalmi overlay háló építése a [Pitsilis-2004] cikk alapján
- Abban a cikkben gyakorlatilag a bizalmi szintet a felhasználók Pearson coefficient értéke adja

[Georgios-2006] Kérés szétterjesztése

- Bizalmi overlay háló építése a [Pitsilis-2004] cikk alapján
- Abban a cikkben gyakorlatilag a bizalmi szintet a felhasználók Pearson coefficient értéke adja
- A megépített bizalmi háló nem fontos (limit alatti) éleinek elhagyása, majd ajánlat keresése:

[Georgios-2006] Kérés szétterjesztése

- Bizalmi overlay háló építése a [Pitsilis-2004] cikk alapján
- Abban a cikkben gyakorlatilag a bizalmi szintet a felhasználók Pearson coefficient értéke adja
- A megépített bizalmi háló nem fontos (limit alatti) éleinek elhagyása, majd ajánlat keresése:
  - ajánlatkérés elindítása a szomszédok felé

[Georgios-2006] Kérés szétterjesztése

- Bizalmi overlay háló építése a [Pitsilis-2004] cikk alapján
- Abban a cikkben gyakorlatilag a bizalmi szintet a felhasználók Pearson coefficient értéke adja
- A megépített bizalmi háló nem fontos (limit alatti) éleinek elhagyása, majd ajánlat keresése:
  - ajánlatkérés elindítása a szomszédok felé
  - a peer hozzáfűzi a listához a bizalmi szintjét és továbbküldi a szomszédoknak

[Georgios-2006] Kérés szétterjesztése

- Bizalmi overlay háló építése a [Pitsilis-2004] cikk alapján
- Abban a cikkben gyakorlatilag a bizalmi szintet a felhasználók Pearson coefficient értéke adja
- A megépített bizalmi háló nem fontos (limit alatti) éleinek elhagyása, majd ajánlat keresése:
  - ajánlatkérés elindítása a szomszédok felé
  - a peer hozzáfűzi a listához a bizalmi szintjét és továbbküldi a szomszédoknak
  - ha a peer rendelkezik szavazattal az adott itemre, akkor visszaküldi a lekérdezést a feladónak

[Georgios-2006] Kérés szétterjesztése

- Bizalmi overlay háló építése a [Pitsilis-2004] cikk alapján
- Abban a cikkben gyakorlatilag a bizalmi szintet a felhasználók Pearson coefficient értéke adja
- A megépített bizalmi háló nem fontos (limit alatti) éleinek elhagyása, majd ajánlat keresése:
  - ajánlatkérés elindítása a szomszédok felé
  - a peer hozzáfűzi a listához a bizalmi szintjét és továbbküldi a szomszédoknak
  - ha a peer rendelkezik szavazattal az adott itemre, akkor visszaküldi a lekérdezést a feladónak
- A beérkező listák bizalmi szintjét hasonlósági mértékké konvertálja, majd kiszámítja a legjobb ajánlást

[Han-2004] DHT

- Az itemek és a hozzá tartozó információk szét vannak osztva a peerek között (DHT - Distributed Hash Table) - bucketek

[Han-2004] DHT

- Az itemek és a hozzá tartozó információk szét vannak osztva a peerek között (DHT - Distributed Hash Table) - bucketek
- Minden egyes item minden egyes előforduló rate-jéhez fel van véve egy bucket, amely tartalmazza azon userek azonosítóit, akik legalább ezzel az értékkel szavaztak az aktuális itemre

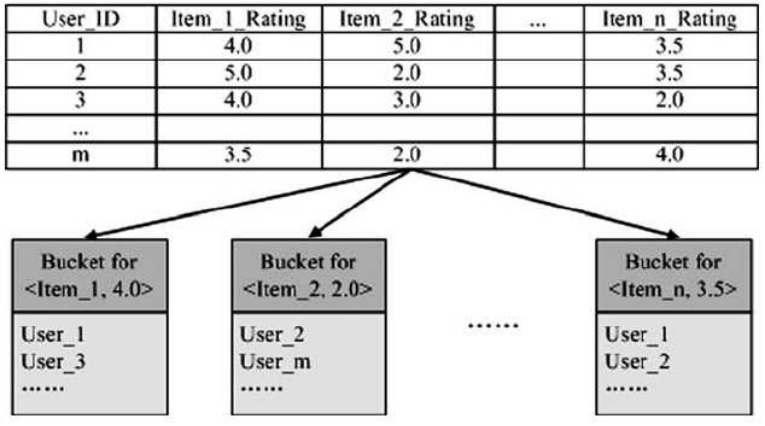


Fig. 1. User database division strategy.

[Han-2004] DHT (2)

- Az ötlet, hogy azon a felhasználóknak hasonló az érdeklődése, akik egy itemre hasonlóan szavaznak

[Han-2004] DHT (2)

- Az ötlet, hogy azon a felhasználóknak hasonló az érdeklődése, akik egy itemre hasonlóan szavaznak
- Az algoritmus két fontos műveletet valósít meg:

[Han-2004] DHT (2)

- Az ötlet, hogy azon a felhasználóknak hasonló az érdeklődése, akik egy itemre hasonlóan szavaznak
- Az algoritmus két fontos műveletet valósít meg:
  - $put(key)$ A DHT overlay konstrukciójára és az itemek elosztására

[Han-2004] DHT (2)

- Az ötlet, hogy azon a felhasználóknak hasonló az érdeklődése, akik egy itemre hasonlóan szavaznak
- Az algoritmus két fontos műveletet valósít meg:
  - $put(key)$ A DHT overlay konstrukciójára és az itemek elosztására
  - $lookup(key)$ A hasonló szavazatokkal rendelkező userek megkeresésére

[Han-2004] DHT (2)

- Az ötlet, hogy azon a felhasználóknak hasonló az érdeklődése, akik egy itemre hasonlóan szavaznak
- Az algoritmus két fontos műveletet valósít meg:
  - $put(key)$ A DHT overlay konstrukciójára és az itemek elosztására
  - $lookup(key)$ A hasonló szavazatokkal rendelkező userek megkeresésére
- A lookup segítségével kialakítható egy lokális train set, ami alapján meg lehet tenni az ajánlást

[Han-2004] DHT (2)

- Az ötlet, hogy azon a felhasználóknak hasonló az érdeklődése, akik egy itemre hasonlóan szavaznak
- Az algoritmus két fontos műveletet valósít meg:
  - $put(key)$ A DHT overlay konstrukciójára és az itemek elosztására
  - $lookup(key)$ A hasonló szavazatokkal rendelkező userek megkeresésére
- A lookup segítségével kialakítható egy lokális train set, ami alapján meg lehet tenni az ajánlást
- Eachmovie adatbázis $\approx$ 0.83 MAE

[Yang-2007] Buddycast

- User based Item rangsoroló módszer:

$$o_u(i) \propto \sum_{\forall u' : u' \in L_i} \log \frac{c(u', u) + \mu \frac{c(u')}{\sum_u c(u')}}{c(u) + \mu}$$

[Yang-2007] Buddycast

- User based Item rangsoroló módszer:

$$o_u(i) \propto \sum_{\forall u' : u' \in L_i} \log \frac{c(u', u) + \mu \frac{c(u')}{\sum_u c(u')}}{c(u) + \mu}$$

  - L_i az i itemre szavazó felhasználók listája

[Yang-2007] Buddycast

- User based Item rangsoroló módszer:

$$o_u(i) \propto \sum_{\forall u' : u' \in L_i} \log \frac{c(u', u) + \mu \frac{c(u')}{\sum_u c(u')}}{c(u) + \mu}$$

  - L_i az i itemre szavazó felhasználók listája
  - $c(u)$ az u felhasználó által szavazott itemek száma

[Yang-2007] Buddycast

- User based Item rangsoroló módszer:

$$o_u(i) \propto \sum_{\forall u' : u' \in L_i} \log \frac{c(u', u) + \mu \frac{c(u')}{\sum_u c(u')}}{c(u) + \mu}$$

  - L_i az i itemre szavazó felhasználók listája
  - $c(u)$ az u felhasználó által szavazott itemek száma
  - $c(u', u)$ az u és u' felhasználók által közösen szavazott itemek száma

[Yang-2007] Buddycast

- User based Item rangsoroló módszer:

$$o_u(i) \propto \sum_{\forall u' : u' \in L_i} \log \frac{c(u', u) + \mu \frac{c(u')}{\sum_u c(u')}}{c(u) + \mu}$$

  - L_i az i itemre szavazó felhasználók listája
  - $c(u)$ az u felhasználó által szavazott itemek száma
  - $c(u', u)$ az u és u' felhasználók által közösen szavazott itemek száma
  - a μ pedig simító paraméter

[Yang-2007] Buddycast (2)

- A leghasonlóbb n felhasználó információinak eltárolása, információcsere (Buddycast algoritmus)

[Yang-2007] Buddycast (2)

- A leghasonlóbb n felhasználó információinak eltárolása, információcsere (Buddycast algoritmus)
- λ valószínűséggel véletlen peer-ek választása, így fenntartva a dinamikát.

[Yang-2007] Buddycast (2)

- A leghasonlóbb n felhasználó információinak eltárolása, információcsere (Buddycast algoritmus)
- λ valószínűséggel véletlen peer-ek választása, így fenntartva a dinamikát.
- User hasonlóság: $c(u', u)$ (közös szavazatok száma)

[Yang-2007] Buddycast (2)

- A leghasonlóbb n felhasználó információinak eltárolása, információcsere (Buddycast algoritmus)
- λ valószínűséggel véletlen peer-ek választása, így fenntartva a dinamikát.
- User hasonlóság: $c(u', u)$ (közös szavazatok száma)
- Csak a hasonló felhsználóknál lévő itemek rangsorolása

[Wang-2006] BuddyTable

- Item based Item rangsoroló módszer:

$$R_{I_a,P_i} \propto \sum_{\forall I_b : I_b \in L_i} \log \frac{2R_{I_b}(I_a)}{R_{I_a}} + \log R_{I_a}$$

[Wang-2006] BuddyTable

- Item based Item rangsoroló módszer:

$$R_{I_a,P_i} \propto \sum_{\forall I_b : I_b \in L_i} \log \frac{2R_{I_b}(I_a)}{R_{I_a}} + \log R_{I_a}$$

  - $R_{I_b}(I_a)$ az a és b item együttelőfordulási valószínűsége

[Wang-2006] BuddyTable

- Item based Item rangsoroló módszer:

$$R_{I_a,P_i} \propto \sum_{\forall I_b : I_b \in L_i} \log \frac{2R_{I_b}(I_a)}{R_{I_a}} + \log R_{I_a}$$

  - $R_{I_b}(I_a)$ az a és b item együttelőfordulási valószínűsége
  - R_{I_a} pedig az a item előfordulási valószínűsége

[Wang-2006] BuddyTable

- Item based Item rangsoroló módszer:

$$R_{I_a,P_i} \propto \sum_{\forall I_b : I_b \in L_i} \log \frac{2R_{I_b}(I_a)}{R_{I_a}} + \log R_{I_a}$$

  - $R_{I_b}(I_a)$ az a és b item együttelőfordulási valószínűsége
  - R_{I_a} pedig az a item előfordulási valószínűsége
- Azon felhasználók letöltési sorában lévő elemek között számolható, akik töltöttek le az i felhasználótól

[Wang-2006] BuddyTable

- Item based Item rangsoroló módszer:

$$R_{I_a,P_i} \propto \sum_{\forall I_b : I_b \in L_i} \log \frac{2R_{I_b}(I_a)}{R_{I_a}} + \log R_{I_a}$$

  - $R_{I_b}(I_a)$ az a és b item együttelőfordulási valószínűsége
  - R_{I_a} pedig az a item előfordulási valószínűsége
- Azon felhasználók letöltési sorában lévő elemek között számolható, akik töltöttek le az i felhasználótól
- Valamint megadják a relevancia lista frissítésének módját (háttérben futó eseményvezérelt program)

[Wang-2006] BuddyTable

- Item based Item rangsoroló módszer:

$$R_{I_a,P_i} \propto \sum_{\forall I_b : I_b \in L_i} \log \frac{2R_{I_b}(I_a)}{R_{I_a}} + \log R_{I_a}$$

  - $R_{I_b}(I_a)$ az a és b item együttelőfordulási valószínűsége
  - R_{I_a} pedig az a item előfordulási valószínűsége
- Azon felhasználók letöltési sorában lévő elemek között számolható, akik töltöttek le az i felhasználótól
- Valamint megadják a relevancia lista frissítésének módját (háttérben futó eseményvezérelt program)
- A topN-User és TopN-Item algoritmusok közt van a teljesítmény

[Castagnos-2007] AURA

- Minden peer rendelkezik egy személyes, publikus, és csoport profillal, valamint négy listával:

[Castagnos-2007] AURA

- Minden peer rendelkezik egy személyes, publikus, és csoport profillal, valamint négy listával:
  - azonos csoportba tartozó felhasználók

[Castagnos-2007] AURA

- Minden peer rendelkezik egy személyes, publikus, és csoport profillal, valamint négy listával:
  - azonos csoportba tartozó felhasználók
  - limit feletti hasonlóságú felhasználók

[Castagnos-2007] AURA

- Minden peer rendelkezik egy személyes, publikus, és csoport profillal, valamint négy listával:
  - azonos csoportba tartozó felhasználók
  - limit feletti hasonlóságú felhasználók
  - fekete lista

[Castagnos-2007] AURA

- Minden peer rendelkezik egy személyes, publikus, és csoport profillal, valamint négy listával:
  - azonos csoportba tartozó felhasználók
  - limit feletti hasonlóságú felhasználók
  - fekete lista
  - aktív felhasználók listája

[Castagnos-2007] AURA

- Minden peer rendelkezik egy személyes, publikus, és csoport profillal, valamint négy listával:
  - azonos csoportba tartozó felhasználók
  - limit feletti hasonlóságú felhasználók
  - fekete lista
  - aktív felhasználók listája
- Az ajánlás az első két lista felhasználói alapján számolódik Pearson hasonlósággal

[Castagnos-2007] AURA

- Minden peer rendelkezik egy személyes, publikus, és csoport profillal, valamint négy listával:
  - azonos csoportba tartozó felhasználók
  - limit feletti hasonlóságú felhasználók
  - fekete lista
  - aktív felhasználók listája
- Az ajánlás az első két lista felhasználói alapján számolódik Pearson hasonlósággal
- Ha a peer profilja változik, akkor azt megosztja az aktív felhasználókkal

[Castagnos-2007] AURA

- Minden peer rendelkezik egy személyes, publikus, és csoport profillal, valamint négy listával:
  - azonos csoportba tartozó felhasználók
  - limit feletti hasonlóságú felhasználók
  - fekete lista
  - aktív felhasználók listája
- Az ajánlás az első két lista felhasználói alapján számolódik Pearson hasonlósággal
- Ha a peer profilja változik, akkor azt megosztja az aktív felhasználókkal
- GroupLens adatbázison $\approx$ 0.85 MAE

Clustering Items for Collaborative Filtering [C

- Az itemek közötti hasonlóságot Pearson correlációval mérték (a negatív hasonlóságot figyelmen kívül hagyták, mert a klaszterező algoritmusaik nem kezelték a negatív hasonlóságot)

Clustering Items for Collaborative Filtering [C

- Az itemek közötti hasonlóságot Pearson correlációval mérték (a negatív hasonlóságot figyelmen kívül hagyták, mert a klaszterező algoritmusaik nem kezelték a negatív hasonlóságot)
- A klaszterező algoritmusok:

Clustering Items for Collaborative Filtering [C

- Az itemek közötti hasonlóságot Pearson correlációval mérték (a negatív hasonlóságot figyelmen kívül hagyták, mert a klaszterező algoritmusaik nem kezelték a negatív hasonlóságot)
- A klaszterező algoritmusok:
  - Average link hierarchical agglomeative

Clustering Items for Collaborative Filtering [C

- Az itemek közötti hasonlóságot Pearson correlációval mérték (a negatív hasonlóságot figyelmen kívül hagyták, mert a klaszterező algoritmusaik nem kezelték a negatív hasonlóságot)
- A klaszterező algoritmusok:
  - Average link hierarchical agglomeative
  - A Robust Clustering Algorithm for Categorical Attributes

Clustering Items for Collaborative Filtering [C

- Az itemek közötti hasonlóságot Pearson correlációval mérték (a negatív hasonlóságot figyelmen kívül hagyták, mert a klaszterező algoritmusaik nem kezelték a negatív hasonlóságot)
- A klaszterező algoritmusok:
  - Average link hierarchical agglomeative
  - A Robust Clustering Algorithm for Categorical Attributes
  - kMetis és hMetis (gráf partícionálók)

Clustering Items for Collaborative Filtering [C

- Az itemek közötti hasonlóságot Pearson correlációval mérték (a negatív hasonlóságot figyelmen kívül hagyták, mert a klaszterező algoritmusaik nem kezelték a negatív hasonlóságot)
- A klaszterező algoritmusok:
  - Average link hierarchical agglomeative
  - A Robust Clustering Algorithm for Categorical Attributes
  - kMetis és hMetis (gráf partícionálók)
- A legjobb eredményt globálisan akkor érték el, amikor nem partícionálták az itemeket

Clustering Items for Collaborative Filtering [C

- Az itemek közötti hasonlóságot Pearson correlációval mérték (a negatív hasonlóságot figyelmen kívül hagyták, mert a klaszterező algoritmusaik nem kezelték a negatív hasonlóságot)
- A klaszterező algoritmusok:
  - Average link hierarchical agglomeative
  - A Robust Clustering Algorithm for Categorical Attributes
  - kMetis és hMetis (gráf partícionálók)
- A legjobb eredményt globálisan akkor érték el, amikor nem partícionálták az itemeket
- Voltak viszont olyan esetek (klaszterek), amikor a kalszteretés jobb eredményt ért el.

An Online Social Network-based Recommend

- Adatbázias: BoardGameGeek (online tábla és kártyajátékok)

An Online Social Network-based Recommend

- Adatbázias: BoardGameGeek (online tábla és kártyajátékok)
- Az ötlet a mátrix fakrotizáció

An Online Social Network-based Recommend

- Adatbázias: BoardGameGeek (online tábla és kártyajátékok)
- Az ötlet a mátrix fakrotizáció
- PMF (probabilistic matrix factorization) algoritmus használata $R \approx UG$

An Online Social Network-based Recommend

- Adatbázias: BoardGameGeek (online tábla és kártyajátékok)
- Az ötlet a mátrix fakrotizáció
- PMF (probabilistic matrix factorization) algoritmus használata $R \approx UG$
- A bizalmi viszonyok leírására az F mátrixot használták $R \approx FUG$

An Online Social Network-based Recommend

- Adatbázias: BoardGameGeek (online tábla és kártyajátékok)
- Az ötlet a mátrix fakrotizáció
- PMF (probabilistic matrix factorization) algoritmus használata $R \approx UG$
- A bizalmi viszonyok leírására az F mátrixot használták $R \approx FUG$
- A megoldandó feladat pedig: $F^{-1}R \approx UG$

An Online Social Network-based Recommend

- Adatbázias: BoardGameGeek (online tábla és kártyajátékok)
- Az ötlet a mátrix fakrotizáció
- PMF (probabilistic matrix factorization) algoritmus használata $R \approx UG$
- A bizalmi viszonyok leírására az F mátrixot használták $R \approx FUG$
- A megoldandó feladat pedig: $F^{-1}R \approx UG$
- A négyzetes helyreállítási hiba 10 iteráció után ≈ 0.4

Influence in Ratings-Based Recommender Sys

- Minden egyes felhasználó rendezi a többi felhasználót a következők szerint:

$$|P_{ja}(M_U) - P_{ja}(M_{U-\{U_i\}})| \geq \delta, \forall j \neq i$$

ahol a $P_{ja}(M_U)$ a predikció az a itemre a u_j usernek az M_U modell alapján

Influence in Ratings-Based Recommender Sys

- Minden egyes felhasználó rendezi a többi felhasználót a következők szerint:

$$|P_{ja}(M_U) - P_{ja}(M_{U-\{U_i\}})| \geq \delta, \forall j \neq i$$

  ahol a $P_{ja}(M_U)$ a predikció az a itemre a u_j usernek az M_U modell alapján

- azaz hányszor megy a predikció egy határ fölé ha az u_i felhasználó nélkül építjük a modellt az u_j felhasználónak

Qulaitative Factors

- Number of ratings

Qulaitative Factors

- Number of ratings
- Degree of agreement with others: $\frac{1}{k}\sum_{a=1}^{k}|R_{ia} - \overline{R}_a|$

Qulaitative Factors

- Number of ratings
- Degree of agreement with others: $\frac{1}{k}\sum_{a=1}^{k}|R_{ia}-\overline{R}_a|$
- Rarity of the rated item: $\frac{1}{k}\sum_{j\in I_{u_i}}\frac{1}{frac(j)}$, ahol I_{u_i} azon itemek halmaza, amelyre az u_i felhasználó szavazott. (IDF jellegű)

Qulaitative Factors

- Number of ratings
- Degree of agreement with others: $\frac{1}{k}\sum_{a=1}^{k}|R_{ia}-\overline{R}_a|$
- Rarity of the rated item: $\frac{1}{k}\sum_{j\in I_{u_i}}\frac{1}{frac(j)}$, ahol I_{u_i} azon itemek halmaza, amelyre az u_i felhasználó szavazott. (IDF jellegű)
- Standard deviation in one's rating: $R_{ik}-\overline{R}_i$

Qulaitative Factors

- Degree of similarity with top neighbors: $u_i : \frac{1}{k} \sum_{j=1}^{k} W_{ij}$

Qulaitative Factors

- Degree of similarity with top neighbors: $u_i : \frac{1}{k} \sum_{j=1}^{k} W_{ij}$
  - magas érték jelenthet nagyobb ráhatást a többiekre

Qulaitative Factors

- Degree of similarity with top neighbors: $u_i : \frac{1}{k} \sum_{j=1}^{k} W_{ij}$
  - magas érték jelenthet nagyobb ráhatást a többiekre
  - az a felhasználó könnyen helyettesíthető, aki sok másikhoz hasonló

Qulaitative Factors

- Degree of similarity with top neighbors: $u_i : \frac{1}{k} \sum_{j=1}^{k} W_{ij}$
  - magas érték jelenthet nagyobb ráhatást a többiekre
  - az a felhasználó könnyen helyettesíthető, aki sok másikhoz hasonló
- Aggregated popularity of the rated items: Ha a rated itemek népszerűség összege elég magas, akkor a felhasználónak nagy az esélye, hogy sok más felhasználóval van metszete.

Qulaitative Factors

- Degree of similarity with top neighbors: $u_i : \frac{1}{k} \sum_{j=1}^{k} W_{ij}$
  - magas érték jelenthet nagyobb ráhatást a többiekre
  - az a felhasználó könnyen helyettesíthető, aki sok másikhoz hasonló
- Aggregated popularity of the rated items: Ha a rated itemek népszerűség összege elég magas, akkor a felhasználónak nagy az esélye, hogy sok más felhasználóval van metszete.
- Aggregated MoviePopularity * Entropy: kiegyenlíti a népszerűséget és a varianciát.

Ötletek a további munkára

- Item-based CF klaszterezéssel ($K \times |I_i|$ méretű távolságmátrik minden i node-ban)

Ötletek a további munkára

- Item-based CF klaszterezéssel ($K \times |I_i|$ méretű távolságmátrik minden i node-ban)
  - P2P K-Means [Datta-2005]

Ötletek a további munkára

- Item-based CF klaszterezéssel ($K \times |I_i|$ méretű távolságmátrik minden i node-ban)
  - P2P K-Means [Datta-2005]
  - Spectral clustering (???)

Ötletek a további munkára

- Item-based CF klaszterezéssel ($K \times |I_i|$ méretű távolságmátrik minden i node-ban)
  - P2P K-Means [Datta-2005]
  - Spectral clustering (???)
- Model-based methods

Ötletek a további munkára

- Item-based CF klaszterezéssel ($K \times |I_i|$ méretű távolságmátrik minden i node-ban)
  - P2P K-Means [Datta-2005]
  - Spectral clustering (???)
- Model-based methods
  - P2P Naive-Bayes [Kowalczyk-2003]

Ötletek a további munkára

- Item-based CF klaszterezéssel ($K \times |I_i|$ méretű távolságmátrik minden i node-ban)
  - P2P K-Means [Datta-2005]
  - Spectral clustering (???)
- Model-based methods
  - P2P Naive-Bayes [Kowalczyk-2003]
  - Decision trees (???)

Ötletek a további munkára

- Item-based CF klaszterezéssel ($K \times |I_i|$ méretű távolságmátrik minden i node-ban)
  - P2P K-Means [Datta-2005]
  - Spectral clustering (???)
- Model-based methods
  - P2P Naive-Bayes [Kowalczyk-2003]
  - Decision trees (???)
  - Bayesian methods (???)

Ötletek a további munkára

- Item-based CF klaszterezéssel ($K \times |I_i|$ méretű távolságmátrik minden i node-ban)
  - P2P K-Means [Datta-2005]
  - Spectral clustering (???)
- Model-based methods
  - P2P Naive-Bayes [Kowalczyk-2003]
  - Decision trees (???)
  - Bayesian methods (???)
  - SVM (???)

Ötletek a további munkára

- Más megközelítés az ajánlásra (propagáció alapú?) (tesztelés?)

Ötletek a további munkára

- Más megközelítés az ajánlásra (propagáció alapú?) (tesztelés?)
- K adaptation a user-user alapú memory-based megközelítéseknél

Ötletek a további munkára

- Más megközelítés az ajánlásra (propagáció alapú?) (tesztelés?)
- K adaptation a user-user alapú memory-based megközelítéseknél
- P2P LSI/PCA négyzetes mátrixokon (alkalmazás: user-user értékelésekből, web gráf?)

Hivatkozások

[Adomavicius-2005] Gediminas Adomavicius and Er Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17:734–749, 2005.

[Aggarwal-1999] Charu C. Aggarwal, Joel L. Wolf, Kun-Lung Wu, and Philip S. Yu. Horting hatches an egg: a new graph-theoretic approach to collaborative filtering. In *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 201–212, San Diego, California, United States, 1999. ACM.

[Bickson-2007] D. Bickson, D. Malkhi, and Lidong Zhou. Peer-to-Peer rating. In *Peer-to-Peer Computing, 2007. P2P 2007. Seventh IEEE International Conference on*, pages 211–218, 2007.

[Billsus-1998] Daniel Billsus and Michael J. Pazzani. Learning collaborative information filters. In *Proc. 15th International Conf. on Machine Learning*, pages 46–54. Morgan Kaufmann, San Francisco, CA, 1998.

[Castagnos-2007] Sylvain Castagnos and Anne Boyer. Modeling preferences in a distributed recommender system. In *User Modeling 2007*, pages 400–404. 2007.

[Connor-2001] Mark Connor and Jon Herlocker. Clustering items for collaborative filtering, 2001.

[Datta-2005] Souptik Datta, Chris Gianella, and Hillol Kargupta. K-means clustering over peer-to-peer networks. In *The 8th International Workshop on High Performance and distributed Mining, SIAM International Conference on Data Mining*, 2005.

[Georgios-2006] Georgios Pitsilis and Lindsay Marshall. A trust-enabled P2P recommender system. In *Enabling Technologies: Infrastructure for Collaborative Enterprises, 2006. WETICE '06. 15th IEEE International Workshops on*, pages 59–64, 2006.

[Goldberg-2001] K. Goldberg, T. Roeder, D. Gupta, and C. Perkins. Eigenstate: A constant time collaborative filtering algorithm. *Information Retrieval*, 4(2):133–151, 2001.

[Han-2004] Peng Han, Bo Xie, Fan Yang, and Ruimin Shen. A scalable P2P recommender system based on dist-

ributed collaborative filtering. *Expert Systems with Applications*, 27(2):203–210, August 2004.

[Hu-2006] Rong Hu and Yansheng Lu. A hybrid user and item-based collaborative filtering with smoothing on sparse data. *International Conference on Artificial Reality and Telexistence*, 0:184–189, 2006.

[Huang-2004] Zan Huang, Daniel Zeng, and Hsinchun Chen. Link analysis for recommendation under sparse data a link analysis approach to recommendation under sparse data. 2004.

[Johnson-2006] Jason K. Johnson, Dmitry M. Malioutov, and Alan S. Willsky. Walk-sum interpretation and analysis of gaussian belief propagation. In Y. Weiss, B. Schölkopf, and J. Platt, editors, *Advances in Neural Information Processing Systems 18*, pages 579–586. MIT Press, Cambridge, MA, 2006.

[Jorge-2008] Jorge Ar, Inmar Givoni, Jeremy H, and Danny Tarlow. An online social network-based recommendation system. 2008.

[Konstan-1999] J. Konstan, A. Borchers, J. Riedl, and J Herlocker. An algorithmic frame-

work for performing collaborative filtering. http://www.grouplens.org/papers/pdf/algs.pdf, 1999.

[Kowalczyk-2003] Wojtek Kowalczyk, Mark Jelasity, and A. E. Eiben. Towards data mining in large and fully distributed peer-to-peer overlay networks. In *In Proceedings of BNAIC03*, pages 203–210, 2003.

[Moallemi-2006] Ciamac Cyrus Moallemi and Benjamin Van Roy. Consensus propagation. *CoRR*, abs/cs/0603078, 2006.

[Moffat-1996] A. Moffat and L. Stuiver. Exploiting clustering in inverted file compression. *Data Compression Conference*, 0:82, 1996.

[Park-2008] Yoon-Joo Park and Alexander Tuzhilin. The long tail of recommender systems and how to leverage it. In *Proceedings of the 2008 ACM conference on Recommender systems*, pages 11–18, Lausanne, Switzerland, 2008. ACM.

[Pitsilis-2004] Georgios Pitsilis, Georgios Pitsilis, Lindsay Marshall, and Lindsay Marshall. A model of trust derivation from evidence for use in recommendation systems. 2004.

[Rashid-2005] Al Mamunur Rashid, George Karypis, and John Riedl. Influence in ratings-based recommender systems: An algorithm-independent approach. In *SDM*, 2005.

[Resnick-1994] Paul Resnick, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom, and John Riedl. Grouplens: An open architecture for collaborative filtering of netnews. pages 175–186. ACM Press, 1994.

[Sarwar-2000] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl. Application of dimensionality reduction in recommender systems–a case study, 2000.

[Sarwar-2001] Badrul Sarwar, George Karypis, Joseph Konstan, and John Reidl. Item-based collaborative filtering recommendation algorithms. pages 285–295, New York, NY, USA, 2001. ACM.

[Tveit-2001] Amund Tveit. Peer-to-peer based recommendations for mobile commerce. In *WMC '01: Proceedings of the 1st international workshop on Mobile commerce*, pages 26–29, New York, NY, USA, 2001. ACM.

[Wang-2006] Jun Wang, Johan Pouwelse, Reginald L. Lagendijk, and Marcel J. T. Reinders. Distributed col-

laborative filtering for peer-to-peer file sharing systems. In *Proceedings of the 2006 ACM symposium on Applied computing*, pages 1026–1030, Dijon, France, 2006. ACM.

[Yang-2007] J. Yang, J. Wang, M. Clements, J. A. Pouwelse, A. P. de Vries, and M. J. T. Reinders. An epidemic-based P2P recommender system. In *Workshop on Large Scale Distributed Systems for Information Retrieval (LSDS-IR)*, pages 11–15, Amsterdam, The Netherlands, 2007. SIGIR 2007 Workshop.

[Yu-2004] Kai Yu, Anton Schwaighofer, Volker Tresp, Xiaowei Xu, and Hans-Peter Kriegel. Probabilistic Memory-Based collaborative filtering. *IEEE Transactions on Knowledge and Data Engineering*, 16(1):56–69, 2004.

[Zhou-2008] Tao Zhou, Zoltan Kuscsik, Jian-Guo Liu, Matus Medo, Joseph R. Wakeling, and Yi-Cheng Zhang. Hybrid algorithms to customize and optimize diversity and accuracy of recommendations. 2008.