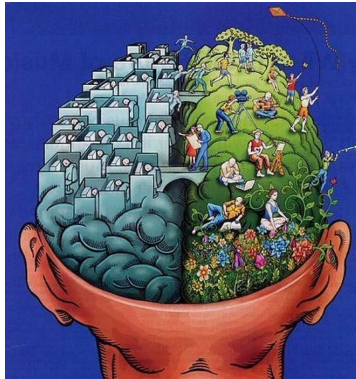




Budapesti Műszaki és Gazdaságtudományi Egyetem Méréstechnika és Információs rendszerek Tanszék



Mesterséges Intelligencia - MI

Ajánló rendszerek

Hullám Gábor – Pataki Béla

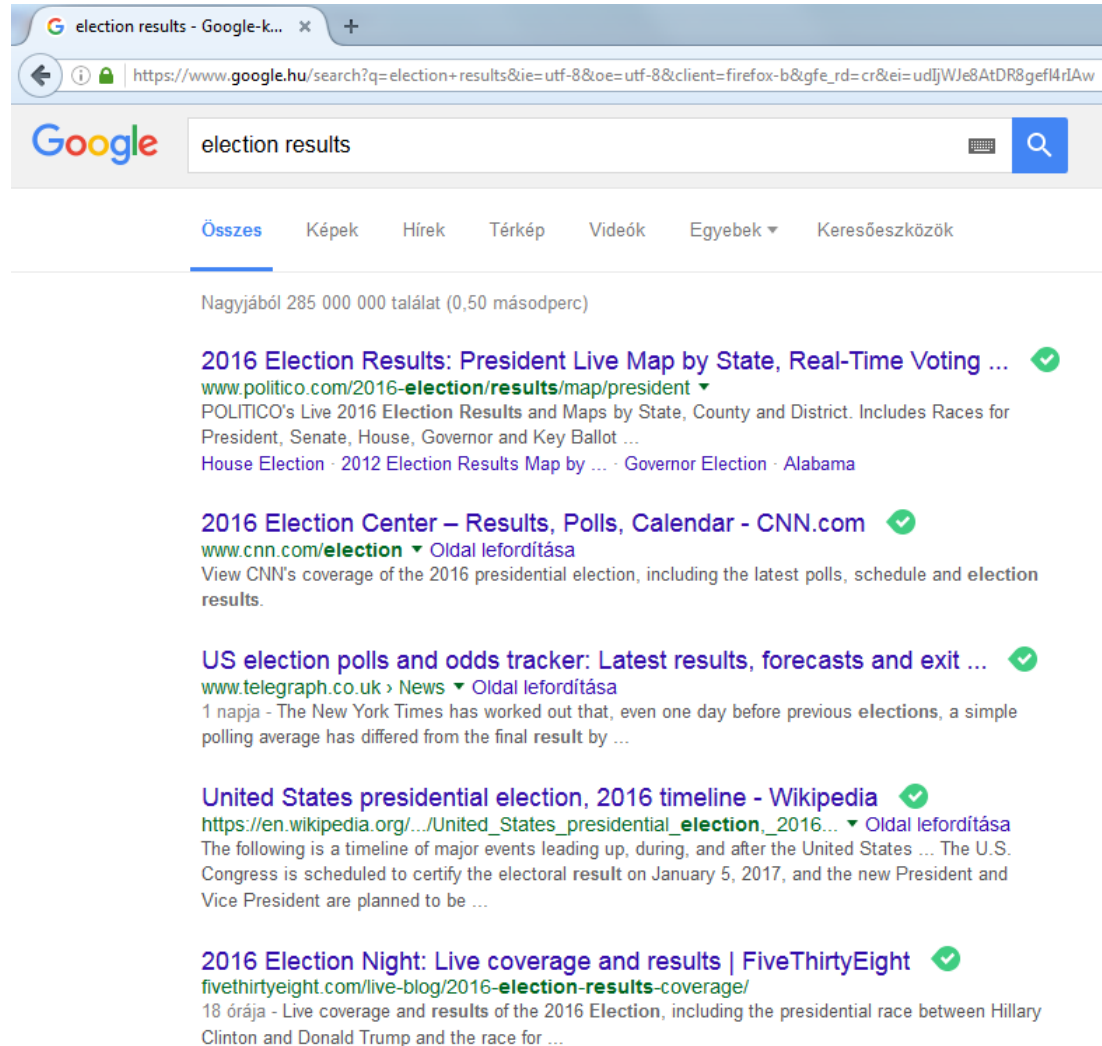


BME I.E. 414, 463-26-79
pataki@mit.bme.hu,
<http://www.mit.bme.hu/general/staff/pataki>



Ajánló rendszerek – hétköznapi példák

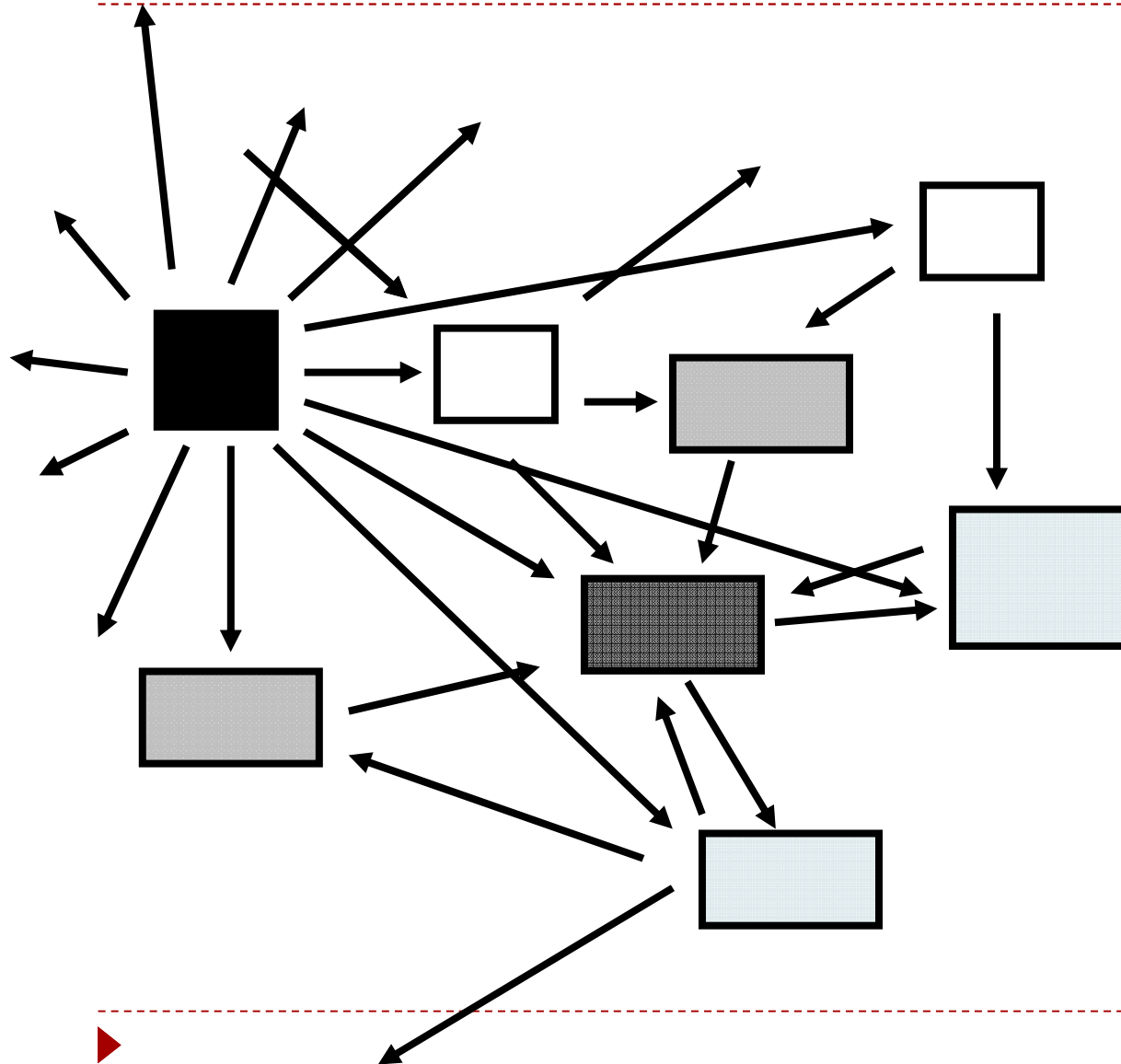
Google



The screenshot shows a Google search interface with the query "election results". The search results page displays several links related to the 2016 US election, each marked with a green checkmark icon. The links include:

- 2016 Election Results: President Live Map by State, Real-Time Voting ...**
www.politico.com/2016-election/results/map/president
POLITICO's Live 2016 Election Results and Maps by State, County and District. Includes Races for President, Senate, House, Governor and Key Ballot ...
House Election · 2012 Election Results Map by ... · Governor Election · Alabama
- 2016 Election Center – Results, Polls, Calendar - CNN.com**
www.cnn.com/election · Oldal lefordítása
View CNN's coverage of the 2016 presidential election, including the latest polls, schedule and election results.
- US election polls and odds tracker: Latest results, forecasts and exit ...**
www.telegraph.co.uk · News · Oldal lefordítása
1 napja - The New York Times has worked out that, even one day before previous elections, a simple polling average has differed from the final result by ...
- United States presidential election, 2016 timeline - Wikipedia**
https://en.wikipedia.org/.../United_States_presidential_election_2016... · Oldal lefordítása
The following is a timeline of major events leading up, during, and after the United States ... The U.S. Congress is scheduled to certify the electoral result on January 5, 2017, and the new President and Vice President are planned to be ...
- 2016 Election Night: Live coverage and results | FiveThirtyEight**
fivethirtyeight.com/live-blog/2016-election-results-coverage/
18 órája - Live coverage and results of the 2016 Election, including the presidential race between Hillary Clinton and Donald Trump and the race for ...

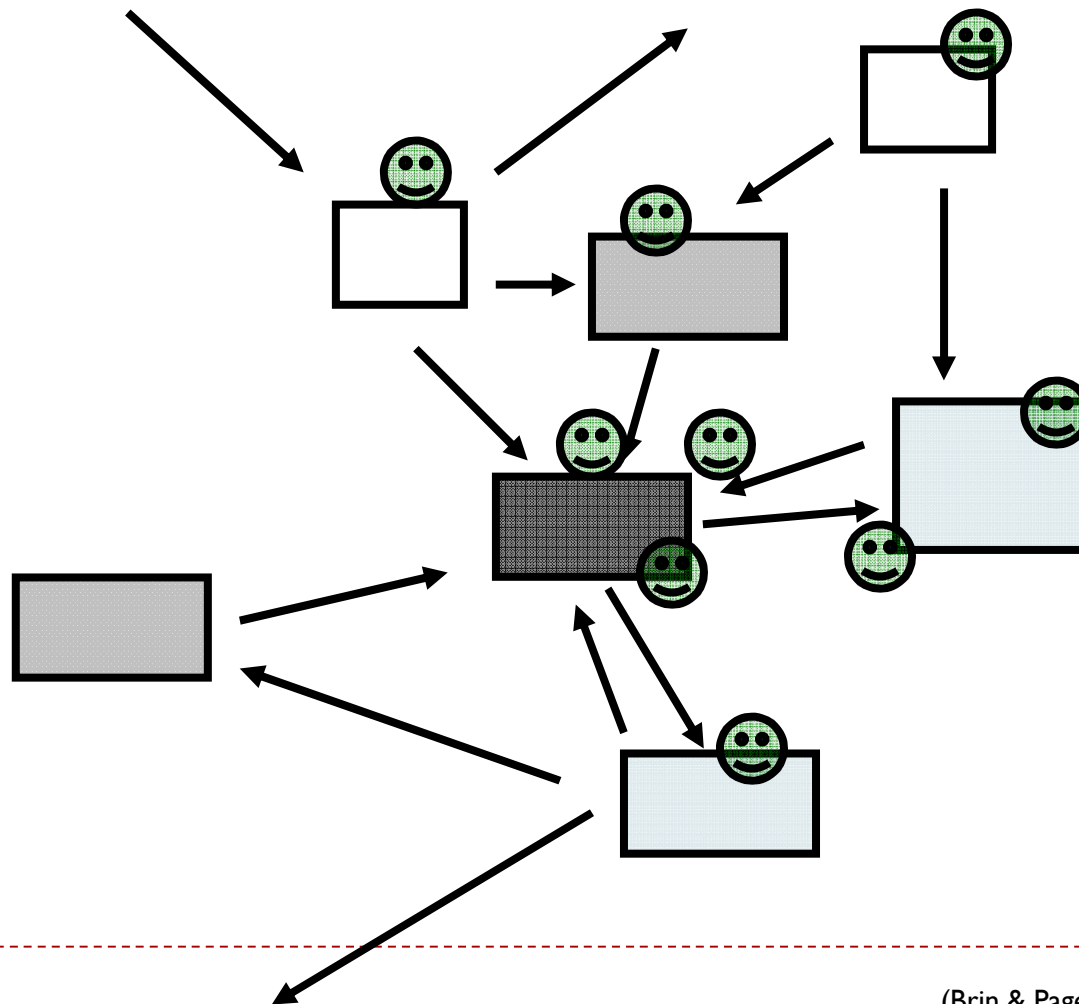
Google PageRank algoritmus



- A bejövő linkek “jó” (pozitív szavazat az oldalra)
- Bejövő linkek egy „jó” oldalról jobbak, mint bejövő linkek egy „rossz” oldalról
- Bejövő linkek olyan oldalakról, ahol sok a kimenő link „nem olyan jó”
- “Jó” és “rossz” meghatározása relatív



Google PageRank algoritmus



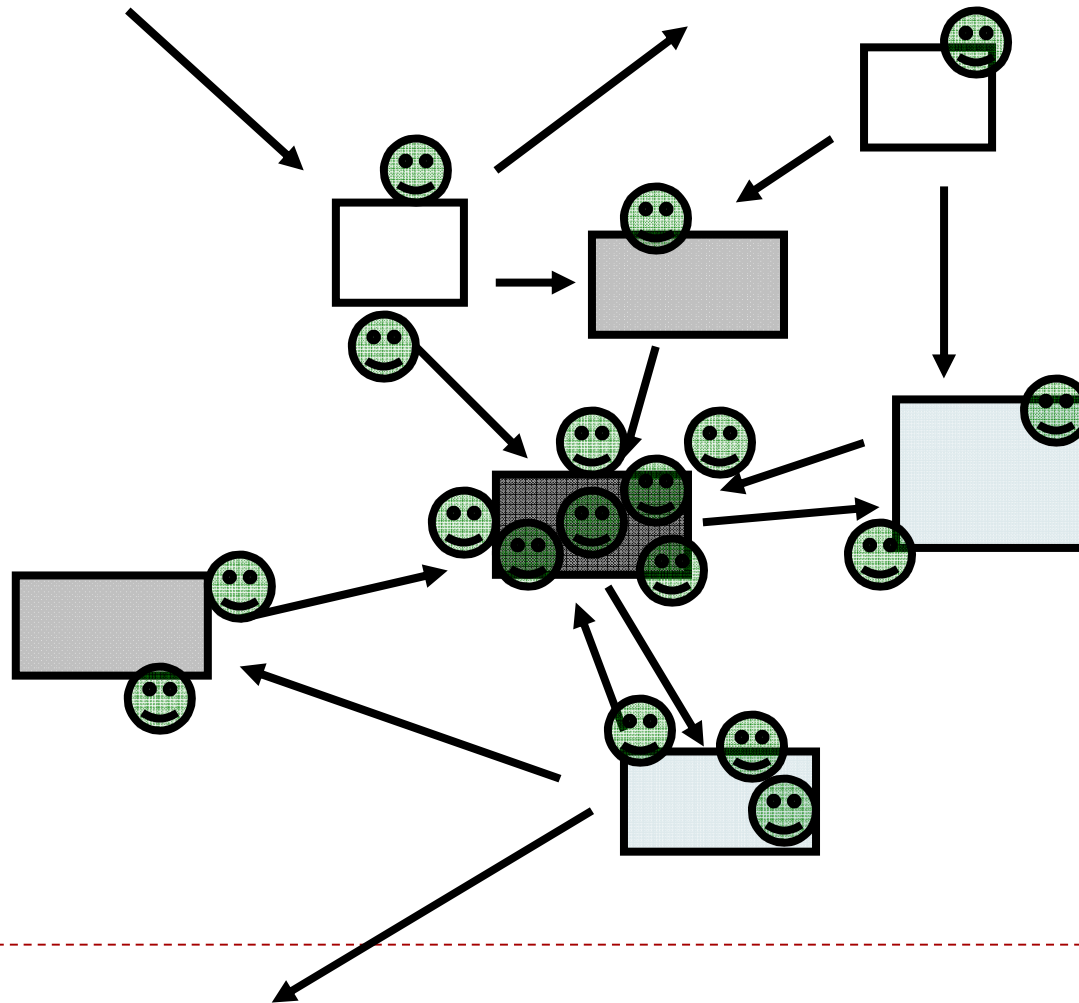
Képzeljünk el egy ágenst,
amelyik

- egy random, adott
oldalon lévő linket követ

VAGY

- egy véletlenszerűen
választott oldalra ugrik

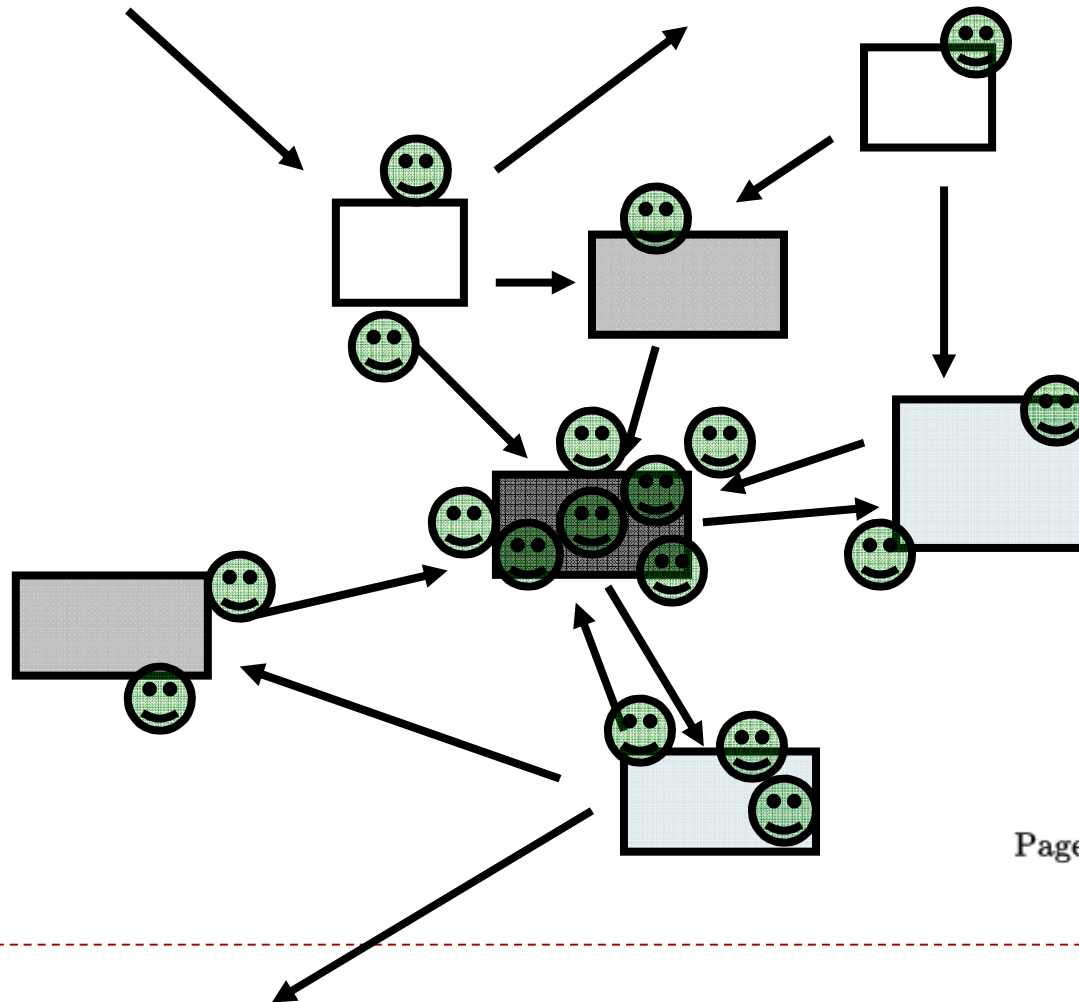
Google PageRank algoritmus



PageRank az alapján rangsorolja az oldalakat, hogy mennyi időt tölt el ott az ágens a bolyongása során

- Vagy, ha több ágens végez bolyongást, akkor a PageRank érték a várható „ágens tömeg” az adott oldalon

Google PageRank algoritmus




- Legyen $j \in M(i)$ azon oldalak halmaza, mely tartalmaz linket i -re
- $L(j)$ az adott j . oldal kimenő linkjeinek a száma
- A „ d ” legyen egy csillapítási tényező
- Az összes oldal száma legyen „ N ”
- Így az i . oldal PageRankje:

$$\text{PageRank}(i) = \frac{1-d}{N} + d \sum_{j \in M(i)} \frac{\text{PageRank}(j)}{L(j)}$$

Ajánló rendszerek – hétköznapi példák

AMAZON



Books

Departments Your Amazon.com Today's Deals Gift Cards & Registry Sell Help

1-12 of 155,767 results for Books : "order"

Show results for

< Any Category

Books

Thrillers & Suspense (1,441)

Suspense (543)

Mystery, Thriller & Suspense (1,878)

Political (137)

Military (95)

Espionage (127)

Literature & Fiction (10,076)

+ See more

Refine by

International Shipping

☐ Ship to Hungary

Amazon Prime

☒ Prime

Eligible for Free Shipping

☐ Free Shipping by Amazon

New Releases

Last 30 days (1,277)

Last 90 days (3,613)

Coming Soon (166)

Author

Vince Flynn

Kyle Mills

Michael A. Stackpole

Walter Jon Williams

Philippa Gregory

James Luceno


James Rollins

Rebecca Cantrell

Sean Williams

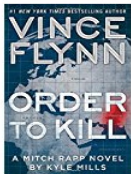
Nicola Barber

Stephanie Rowe



Your Orders – View, Modify, Cancel or Track an Order

[Recent Orders](#) | [Open Orders](#) | [Digital Orders](#)




Order to Kill: A Novel (A Mitch Rapp Novel)

Oct 11, 2016

by Vince Flynn and Kyle Mills

Hardcover

\$17.39 ~~\$28.99~~ 

Get it by **Friday, Nov 11**

More Buying Choices

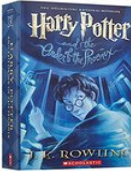
\$5.97 used & new (94 offers)

Kindle Edition

\$19.04

Whispersync for Voice-ready

Other Formats: [Paperback](#), [Audiobook](#), [Mass Market Paperback](#), [Audio CD](#)



Harry Potter And The Order Of The Phoenix

Aug 10, 2004

by J.K. Rowling

Paperback

\$7.86 ~~\$12.99~~ 

Get it by **Friday, Nov 11**

More Buying Choices

\$1.98 used & new (441 offers)

Kindle Edition


\$0.00 [kindleunlimited](#)

Read this and over 1 million books with [Kindle Unlimited](#).

\$10.12 to buy

Auto-delivered wirelessly

Other Formats: [Hardcover](#), [Audiobook](#), [Mass Market Paperback](#), [Audio CD](#)



Elli: Mail Order Brides of Trace Hollow Book 1

Nov 4, 2016

by CJ Samuels

Kindle Edition

\$0.00 [kindleunlimited](#)

Read this and over 1 million books with [Kindle Unlimited](#).

\$1.26 to buy

Auto-delivered wirelessly

Ajánló rendszerek – hétköznapi példák

Youtube

YouTube HU Keresés

FROM THE LABS

0:17 / 8:22

From the Labs: Winning the Netflix Prize

12 600 megtekintés

98 4 MEGOSZTÁS

AT&T Tech Channel
Közzététel: 2010. jan. 28.

AT&T Labs researchers and million dollar Netflix Prize co-winners, Chris Volinsky and Robert Bell, describe their three-year quest to improve the collaborative filtering algorithm Netflix relies on to make millions of movie recommendations.

Learn more about the AT&T Tech Channel at <http://techchannel.att.com>

FELIRATKOZÁS

Követhető

AUTOMATIKUS LEJÁTSZÁS

Advances in Ensemble Learning from the Netflix Prize
San Francisco Bay ACM
2,5 E megtekintés
1:17:33

Computing the Singular Value Decomposition | MIT 18.06SC
MIT OpenCourseWare
150 E megtekintés
11:36

Dr. Yehuda Koren - The Netflix Prize Quest
Rutgers
2,5 E megtekintés
1:00:12

Collaborative Filtering | Stanford University
Video Tutorials - All in One
29 E megtekintés
20:53

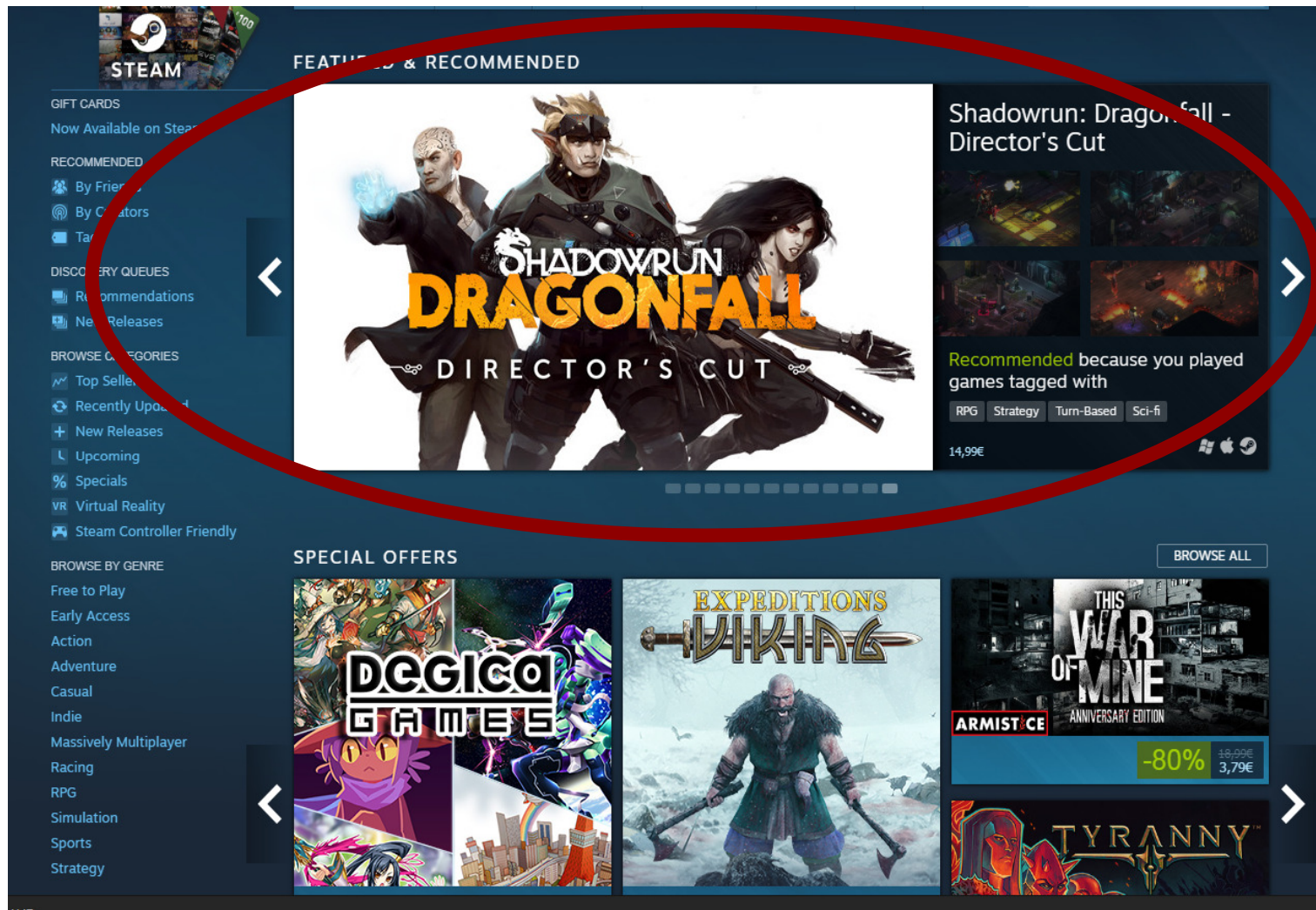
ECS
Michael Wood
50

8 Recommender Systems - Machine Learning Class 10-701
Alex Smola
Neked ajánlott
1:16:26

Recent trends in recommender systems
DATA SCIENCE SUMMIT EUROPE 2
Neked ajánlott
31:45

Ajánló rendszerek – hétköznapi példák

STEAM



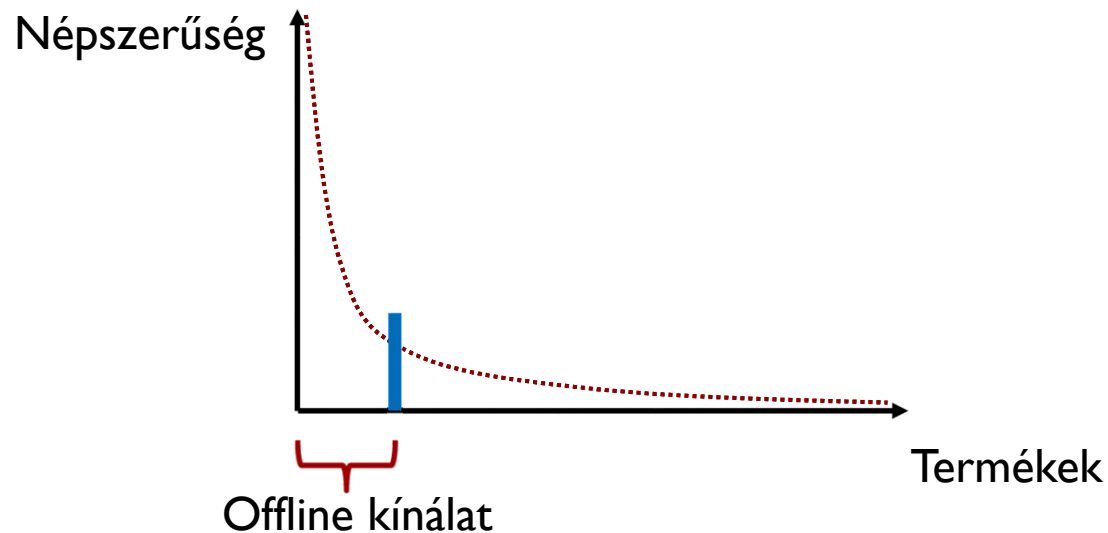


NetFlix kihívás

- ▶ Kitűzött díj: 1'000'000 \$
- ▶ Feltétel: Legyen 10%-kal jobb, mint a NetFlix saját rendszere (CineMatch RMSE: 0.95)
- ▶ Ajánlások: 1-5 *, 17'000 film
- ▶ Nyertes: UV dekompozíciós algoritmus
 - ▶ Michael Harris, Jeffrey Wang, and David Kamm, 2009
 - ▶ Végző változat: több algoritmus együttese

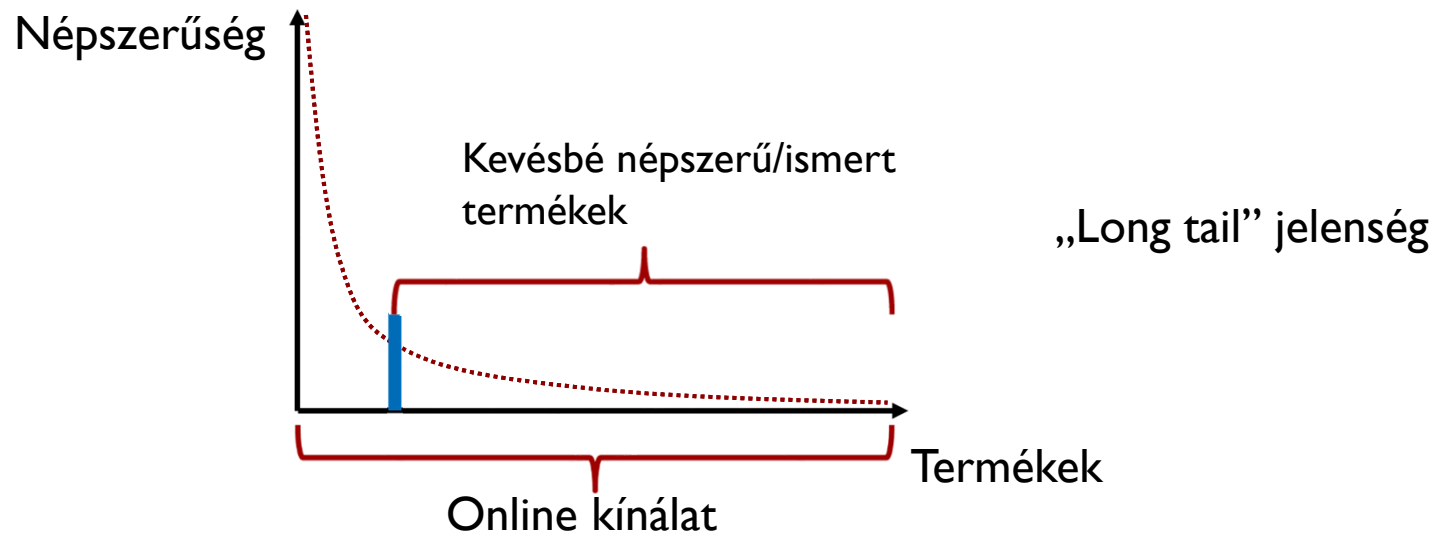
Miért szükségesek az ajánló rendszerek?

- ▶ A fizikai valóságban létező (offline) kereskedelmi egységek, boltok eleve egy szűrt kínálatot tárnak a vásárlók elé
 - ▶ Korlátos mennyiségű árut képesek bemutatni
 - ▶ Népszerű vagy várhatóan népszerű termékeket ajánl
 - ▶ Kritikusok/szakértők véleménye alapján
 - ▶ Korábbi tapasztalatok alapján



Miért szükségesek az ajánló rendszerek?

- ▶ Online térben létező kereskedelmi egységeknél nincs ilyen korlát
 - ▶ ‘Tetszőleges’ mennyiségű árut képesek bemutatni
- ▶ Van viszont egy másik korlát:
 - ▶ A felhasználó korlátos mennyiségű árut képes átlátni

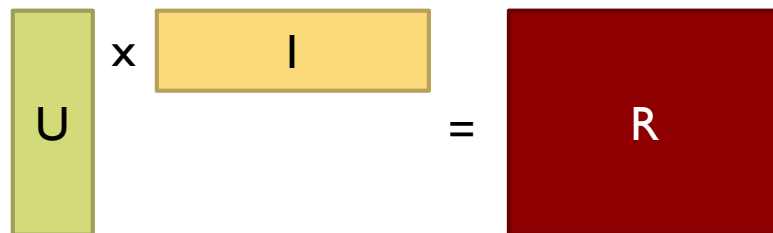


Ajánlási feladat

Adott:

- ▶ U - felhasználók (users),
- ▶ I - termékek (items),
- ▶ $p: U \rightarrow S_u$ - felhasználói profilok
- ▶ $c: I \rightarrow S_i$ - termék leírások
- ▶ $f: U \times I \rightarrow R$ - preferencia (rating) mátrix

Lényeges: nagy mértékű a
hiányzó elemek száma R-ben!



Felhasználó

Termék

	A	B	C	D
User1	5		3	1
User2		4		
User3			2	
User4	2			5

Ajánlási feladat

Felhasználó	Termék				
	A	B	C	D	
	User1	5		3	1
	User2		4		
	User3			2	
	User4	2			5

Cél:

Határozzuk meg azt az $f^*: U \times I \rightarrow R$ leképezést,

- ▶ ami a legjobban közelíti f -et, és
- ▶ teljesen definiált a teljes $U \times I$ téren

Ajánló rendszer – virtuális közösség

- ▶ Virtuális közösség : Olyan emberek egy csoportja, mely
 - ▶ személyre szabott információkat oszt meg / vesz igénybe járulékos kommunikációs költségek nélkül
 - ▶ képesek egymást befolyásolni valós interakció nélkül

W. Hill et al.: „*Recommending And Evaluating Choices In A Virtual Community Of Use.*”, Bellcore; CHI 1995

- ▶ Az ajánlások (legalább részben) emberektől származnak
- ▶ Az ajánlás mellé mellékelni kellene egy bizonyossági mértéket
- ▶ Nem csak egyedi felhasználók számára, hanem felhasználók csoportjai számára is legyen képes ajánlatot adni

Ajánló rendszerek – további szempontok

- ▶ Általában nincs szükség az összes hiányzó elem jóslására
- ▶ Nem kell az összes terméket rangsorolni egy felhasználó szempontjából, elég ha egy kellően nagy eredményhalmazt kap eredményül
- ▶ Nem kell megadni a felhasználónak az összes az magas rangú találatot

		Termék			
		A	B	C	D
Felhasználó	User1	5		3	1
	User2		4		
	User3			2	
	User4	2			5

Honnan származik az értékelés?

- ▶ Felhasználók explicit értékelései alapján
 - ▶ Termékek értékelése (1-5*), vélemények, kedvelések (0/1)
 - ▶ Sokan megtagadják az értékelést
 - ▶ Torzíthat, ezért normalizációra lehet szükség (átlagos értékeléssel)
- ▶ Felhasználók viselkedése alapján
 - ▶ Milyen termékeket vásárolt?
 - ▶ Milyen termékeket nézett meg?
 - ▶ Hol időzött sokáig?

Ajánló rendszerek csoportosítása

Felhasznált információ alapján:

- ▶ **Tartalom alapú** (content-based) módszerek: a felhasználó korábbi értékelései alapján
- ▶ **Kollaboratív szűrés** (collaborative filtering): az ajánlás inputjához hasonló adatok is felhasználásra kerülnek
 - ▶ user-user: a felhasználóhoz hasonló felhasználók szavazatainak felhasználásával
 - ▶ item-item: a cél termékhez hasonló termékek értékelései használatával
- ▶ **Hibrid módszerek**: az előzőek együttes alkalmazása

Ajánló rendszerek csoportosítása

Megközelítés alapján:

- ▶ memória alapú módszerek
 - ▶ hasonlósági metrikák
- ▶ dimenzió redukciós módszerek
 - ▶ mátrix faktorizáció
- ▶ gépi tanulás alapú módszerek
 - ▶ ajánló rendszer, mint osztályozási feladat
- ▶ valószínűségi modell alapú megközelítések

Ajánló rendszerek – profilok készítése

Tartalom alapú rendszerek

- ▶ Termék profilok
 - ▶ Termék tulajdonságok
 - ▶ Tartalom alapú hasonlósági metrikák

Tartalom alapú kollaboratív szűrés

- ▶ Felhasználói profilok
 - ▶ Kedvelt termékek
 - ▶ Viselkedési mutatók
- ▶ Felhasználó és/vagy termék klaszterezés
- ▶ Felhasználók modellezése
 - ▶ Osztályozás

User alapú CF

BellCore: MovieRecommender

- ▶ A résztvevők emailben kaptak egy 500 filmet tartalmazó listát (250 random, 250 populáris) értékelés céljából (1-10 közötti skálán)
 - ▶ Csak egy részhalmazt kell értékelniük
- ▶ Az új résztvevő U^* beküldi az értékelését
- ▶ A keretrendszer összehasonlítja U^* értékelését más (random) felhasználókéval
- ▶ A **leghasznalóbb felhasználók** értékelései alapján a nem értékelt filmek is értékelést kapnak
 - ▶ *Empirical Analysis of Predictive Algorithms for Collaborative Filtering*
Breese, Heckerman, Kadie, UAI98
- ▶ A keretrendszer visszaküld egy emailt az új ajánlásokkal



BellCore: MovieRecommender

- ▶ Kiértékelés
 - ▶ Minden felhasználó értékeléseinek 10%-a legyen a teszhalmaz része
 - ▶ Vizsgáljuk meg a prediktált értékelések és a valódi értékelések közötti korrelációt



Memória alapú CF

- ▶ $v_{i,j}$ = i felhasználó értékelése (szavazata) j termékről
- ▶ I_i = mindazon termékek, melyekről i felhasználó értékelést adott
- ▶ Átlagos értékelés i felhasználónál: $\bar{v}_i = \frac{1}{|I_i|} \sum_{j \in I_i} v_{i,j}$
- ▶ Prediktált értékelés a „aktív felhasználóra” egy súlyozott összeg formájában áll elő

$$p_{a,j} = \bar{v}_a + \kappa \sum_{i=1}^n \underbrace{w(a,i)}_{n \text{ hasonló felhasználó súlya}} (v_{i,j} - \bar{v}_i)$$

Normalizációs tényező

κ hasonló felhasználó súlya



(Breese et al, UAI98)

Memória alapú CF – hasonlóság metrikák

- ▶ K- legközelebbi szomszéd (k-nearest neighbor)

$$w(a, i) = \begin{cases} 1 & \text{if } i \in \text{neighbors}(a) \\ 0 & \text{else} \end{cases}$$

- ▶ Pearson korrelációs koefficiens

$$w(a, i) = \frac{\sum_j (v_{a,j} - \bar{v}_a)(v_{i,j} - \bar{v}_i)}{\sqrt{\sum_j (v_{a,j} - \bar{v}_a)^2 \sum_j (v_{i,j} - \bar{v}_i)^2}}$$

- ▶ Koszinusz távolság:

$$w(a, i) = \sum_j \frac{v_{a,j}}{\sqrt{\sum_{k \in I_a} v_{a,k}^2}} \frac{v_{i,j}}{\sqrt{\sum_{k \in I_i} v_{i,k}^2}}$$



Tartalom alapú ajánló rendszerek

LIBRA Book Recommender

- CF módszerek felteszik, hogy egyes felhasználók preferenciái általában egyeznek más felhasználók preferenciáival.
- Azokat a termékeket, melyeket kevesen értékelték, jellemzően nem javasolja a CF.
- Egyes könyveket kevesen használnak és így kevesen értékelnek
- A CF jól ismert címeket ajánl homogenitást feltételezve az ízlésben
- Mindezek miatt szükség van egy tartalom alapú megközelítésre

Content-Based Book Recommending Using Learning for Text Categorization. Raymond J. Mooney, Lorienne Roy, Univ Texas/Austin; DL-2000



LIBRA Book Recommender - jellemzők

- ▶ Adatbázis szöveges leírással + meta-információ a könyvekről (pl.: Amazon.com-ról)
 - ▶ cím, szerzők, kivonat, könyvkritika, vásárlói hozzászólások, **kapcsolódó szerzők, kapcsolódó címek és kulcsszavak**
- ▶ Felhasználó 1-10 között értékeli a tanítás alapját képző könyveket
- ▶ A rendszer egy (naiv Bayes-háló alapú) modellt készít a felhasználóról
 - ▶ A modell alapján prediktálható: $P(\text{user rating} > 5 | \text{book})$
- ▶ A rendszer informatív jegyek (tulajdonságok) segítségével magyarázatot ad az értékelésekre

Különbségek a MovieRecommender-hez képest:

- Az ajánlás az értékelt terméken alapul és nem más felhasználók választásain
- A memória alapú módszerekkel összehasonlítva a LIBRA konkrét modellt készít a felhasználóról (különböző szavakhoz rendelt súlyok formájában)



LIBRA Book Recommender - példa

The Fabric of Reality:
The Science of Parallel Universes- And Its Implications
by David Deutsch recommended because:

Slot	Word	Strength
DESCRIPTION	MULTIVERSE	75.12
DESCRIPTION	UNIVERSES	25.08
DESCRIPTION	REALITY	22.96
DESCRIPTION	UNIVERSE	15.55
DESCRIPTION	QUANTUM	14.54
DESCRIPTION	INTELLECT	13.86
DESCRIPTION	OKAY	13.75
DESCRIPTION	RESERVATIONS	11.56

The word UNIVERSES is positive due to your ratings:

Title	Rating	Count
<i>The Life of the Cosmos</i>	10	15
<i>Before the Beginning : Our Universe and Others</i>	8	7
<i>Unveiling the Edge of Time</i>	10	3
<i>Black Holes : A Traveler's Guide</i>	9	3
<i>The Inflationary Universe</i>	9	2



Hibrid ajánló rendszerek

Tartalom + CF

Műfaji besorolás

Felhasználói profil

		Airplane	Matrix	Room with a View	...	Hidalgo
		comedy	action	romance	...	action
Joe	27,M,70k	9	7	2		7
Carol	53,F,20k	8		9		
...						
Kumar	25,M,22k	9	3			6
U_a	48,M,81k	4	7	?	?	?



Megközelítés osztályozási feladatként

Osztályozási feladat: (felhasználó,film) párok értékelésének leképzése {likes,dislikes} címkékre

Tanító adat: ismert kedvelések (likes/dislikes)

Teszt adat: aktív felhasználók

Jegyek (features): (felhasználó,film) párok tulajdonságai

		Airplane	Matrix	Room with a View	...	Hidalgo
		comedy	action	romance	...	action
Joe	27,M,70k	1	1	0		1
Carol	53,F,20k	1		1		0
...						
Kumar	25,M,22k	1	0	0		1
U_a	48,M,81k	0	1	?	?	?

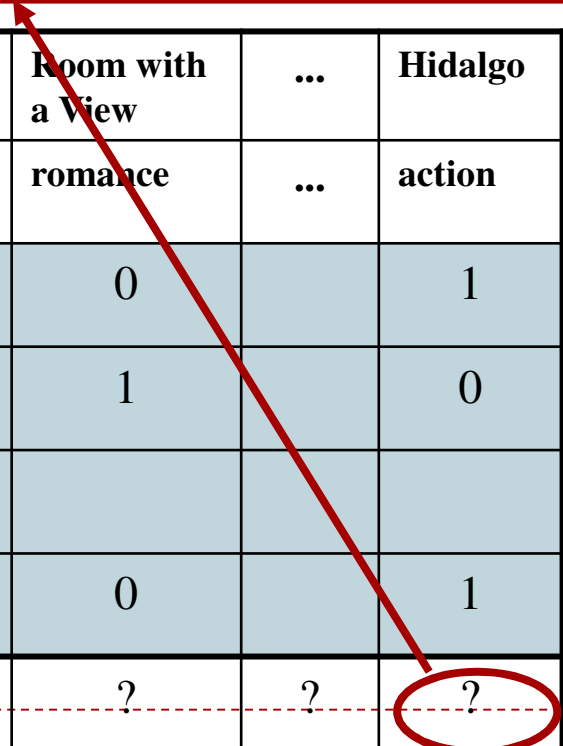


Hibrid ajánló rendszerek

genre={romance}, age=48, sex=male, income=81k,
usersWhoLikedMovie={Carol}, moviesLikedByUser={Matrix,Airplane}, ...

genre={action}, age=48, sex=male, income=81k, usersWhoLikedMovie =
{Joe,Kumar}, moviesLikedByUser={Matrix,Airplane},...

		Airplane	Matrix	Room with a View	...	Hidalgo
		comedy	action	romance	...	action
Joe	27,M,70k	1	1	0		1
Carol	53,F,20k	1		1		0
...						
Kumar	25,M,22k	1	0	0		1
U_a	48,M,81k	1	1	?	?	?



Osztályozási szabályok tanulása

genre={romance}, age=48, sex=male, income=81k,
usersWhoLikedMovie={Carol}, moviesLikedByUser={Matrix,Airplane}, ...

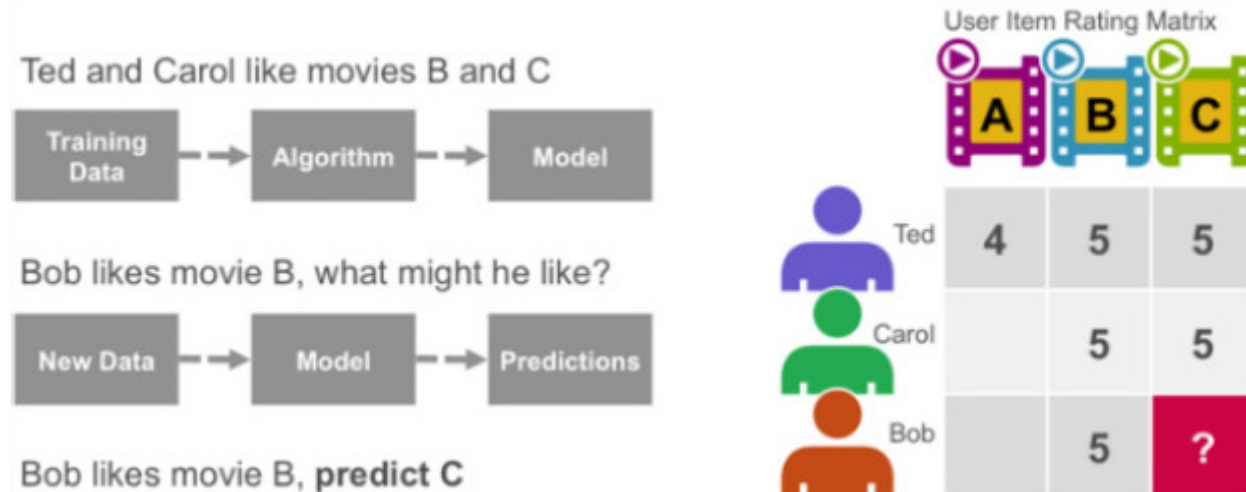
genre={action}, age=48, sex=male, income=81k, usersWhoLikedMovie =
{Joe,Kumar}, moviesLikedByUser={Matrix,Airplane},...

- Osztályozás megvalósítása : szabálytanulással (RIPPER[#])
 - If *NakedGun33/13* in *moviesLikedByUser* and *Joe* in *usersWhoLikedMovie* and *genre=comedy* then predict *likes(U,M)*
 - If *age>12* and *age<17* and *HolyGrail* in *moviesLikedByUser* and *director=MelBrooks* then predict *likes(U,M)*
 - If *Ishtar* in *moviesLikedByUser* then predict *likes(U,M)*
- A memória alapú módszerekkel szemben a szabálytanuló algoritmus modellt készít: hogyan kapcsolódnak a felhasználó preferenciái a termékekhez és más felhasználók preferenciáihoz



Ajánló rendszerek – Mátrix faktorizáció

- Preferencia információ sok felhasználótól
- Feltételezés: akik hasonló dolgokat kedveltek a múltban, a jövőben is így tesznek majd
- Vannak elemek, melyekről van preferencia megadva
- Sok esetben nincs megadva



<https://www.mapr.com/blog/apache-spark-machine-learning-tutorial>

Ajánló rendszerek – Mátrix faktORIZÁCIÓ

- Kérdés: Hogyan döntenének a felhasználók a még nem értékelt elemekről?
- Ez kellene ahhoz, hogy ez alapján másoknak ajánlásokat adhassunk
- Mátrix alapú reprezentáció (ez sokszor hiányos)
- Cél: mátrix „kitöltése”
- Módszer: mátrix faktORIZÁCIÓ

	I1	I2	I3	I4
U1	4	2	-	2
U2	4	1	-	-
U3	-	3	5	2
U4	2	5	-	3
U5	-	-	1	1



Mátrix faktorizáció - 1

Mátrix faktorizáció alkalmazása

- Feltételezés a háttérben léteznek nem ismert jegyek **K** (features)
- jegyek alapján értékelnek a felhasználók: **U**
- jegyek száma jóval kisebb a felhasználókénál : **K** << **U**
- elemek (pl.: filmek): **I**
- Preferencia információ (rating mátrix) : **R**

	I1	I2	I3	I4
U1	4	2	-	2
U2	4	1	-	-
U3	-	3	5	2
U4	2	5	-	3
U5	-	-	1	1



Mátrix faktorizáció -2

A cél megtalálni 2 mátrixot:

- $P = U \times K$ ~ felhasználók és jegyek közötti asszociáció
- $Q = I \times K$ ~ elemek és jegyek közötti asszociáció

melyekre igaz, hogy:

$$P \times Q = R^{\wedge} \quad (R \text{ közelítése})$$

Az u_i felhasználó által a i_j elemre adott rating : $r_{i,j}$

$$\hat{r}_{ij} = p_i^T q_j = \sum_{k=1}^K p_{ik} q_{kj}$$

P és Q „megtalálásához” két (valamilyen módon inicializált) mátrix szorzatát kell összevetni az eredeti R mátrixszal és törekedni iteratívan az eltérés csökkenésére.

Eltérés:

$$e_{ij}^2 = (r_{ij} - \hat{r}_{ij})^2 = (r_{ij} - \sum_{k=1}^K p_{ik} q_{kj})^2$$



Mátrix faktorizáció -3

- ▶ Gradiens számítás p_{ik} és q_{kj} módosításához

$$\begin{aligned}\frac{\partial}{\partial p_{ik}} e_{ij}^2 &= -2(r_{ij} - \hat{r}_{ij})(q_{kj}) = -2e_{ij}q_{kj} \\ \frac{\partial}{\partial q_{kj}} e_{ij}^2 &= -2(r_{ij} - \hat{r}_{ij})(p_{ik}) = -2e_{ij}p_{ik}\end{aligned}$$

- ▶ p_{ik} és q_{kj} módosítása a gradiens alapján

$$\begin{aligned}p'_{ik} &= p_{ik} + \alpha \frac{\partial}{\partial p_{ik}} e_{ij}^2 = p_{ik} + 2\alpha e_{ij}q_{kj} \\ q'_{kj} &= q_{kj} + \alpha \frac{\partial}{\partial q_{kj}} e_{ij}^2 = q_{kj} + 2\alpha e_{ij}p_{ik}\end{aligned}$$

- ▶ Leállni akkor célszerű, amikor a teljes hiba (E) minimális

$$E = \sum_{(u_i, l_j, r_{ij}) \in T} e_{ij} = \sum_{(u_i, l_j, r_{ij}) \in T} \left(r_{ij} - \sum_{k=1}^K p_{ik} q_{kj} \right)^2$$



Mátrix faktorizáció -4

- ▶ Találtunk egy olyan P és Q mátrixot, ami jól közelíti R ismert részeit
- ▶ A nem ismert részeire pedig ad egy becslést
- ▶ Így már R használható ajánláshoz



Budapest University of Technology and Economics
Department of Measurement and Information Systems

