



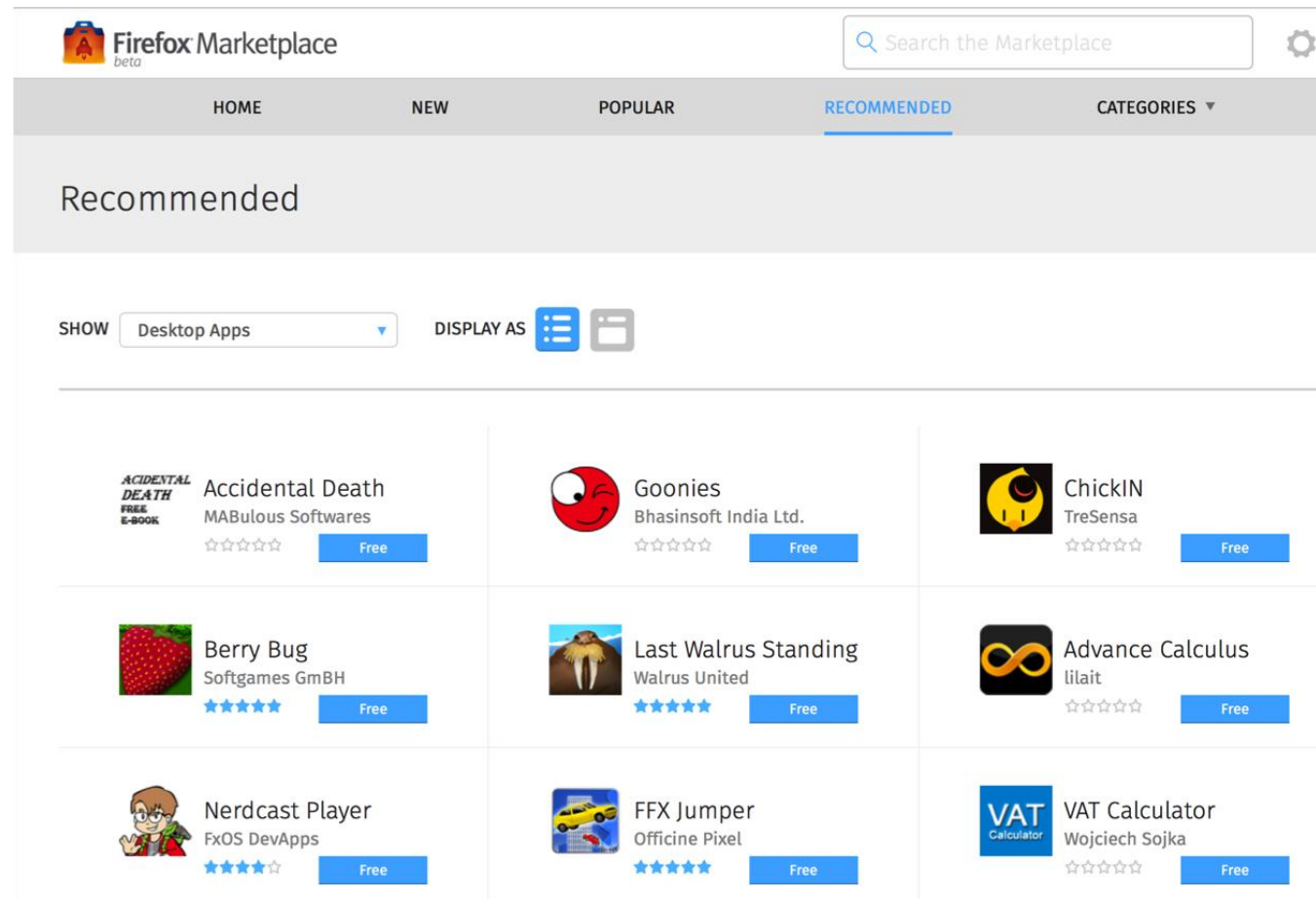
Mašinsko učenje

Sistemi za preporuku

Nemanja Maček

- Uvodne napomene
- Šta je sistem preporuke?
- Pristupi
- Kolaborativno filtriranje
- Preporuke zasnovne na sadržaju
- Zaključne napomene

Mozilla FirefoxOS Marketplace.



Od pretrage ka preporukama.

- „The Web is leaving the era of search and entering one of discovery. What's the difference? Search is what you do when you're looking for something. Discovery is when something wonderful that you didn't know existed, or didn't know how to ask for, finds you.“ *

* Izvor: CNN Money – „The race to create a smart Google.“

Vrednost preporuka.

- Netflix: 2/3 filmova je odgledano na osnovu preporuka.
- Google News: preporuke generišu 38% više pregleda.
- Amazon: 35% prodaje potiče od preporuka.
- Choicestream: 28% ljudi bi kupovalo više muzike ukoliko bi našli ono što im se sviđa.

* Podaci se odnose na 2015. godinu

Problem preporuka (engl. *recommender problem*).

- Problem preporuka je procena funkcije koristi (engl. *utility function*) koja će predvideti kako će se korisniku artikal dopasti (npr. film, muzički album, knjiga itd.)
- Neka je C skup korisnika, a S skup artikala za preporuku.
- Funkcija koja meri korisnost artikla s za korisnika c data je sa $u: C \times S \rightarrow R$, gde je R skup preporučenih artikala.
- Za svakog korisnika c potrebno je odabrati article $s_1, s_2, \dots, s_i$ koji maksimizuju u :

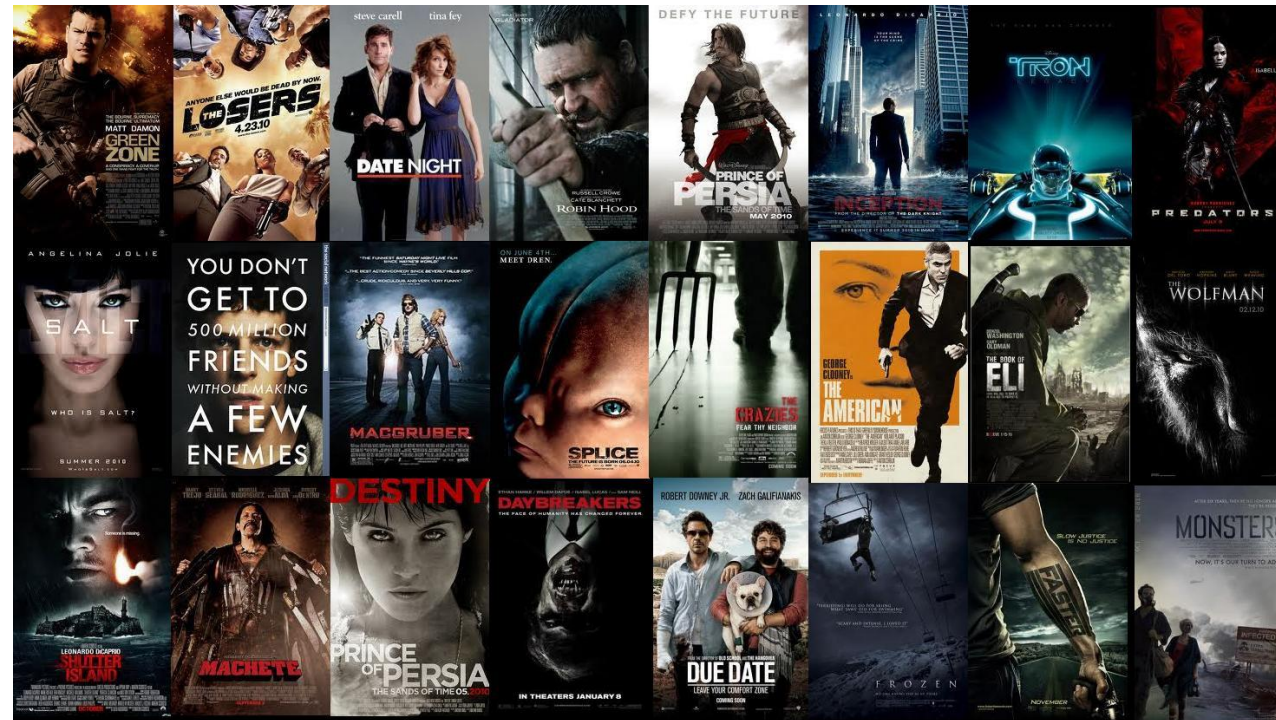
$$s'_c = \operatorname{argmax}_{s \in S} u(c, s)$$

Dobra preporuka.

- Dobra preporuka je relevantna za korisnika, odnosno personalizovana.
- Dobra preporuka je raznolika, odnosno predstavlja sve moguće interese jednog korisnika.
- Korisniku se ne preporučuju artikli po drugi put (na primer, film koji je već pogledao), ili artikli koje bi u svakom slučaju našao. Drugim rečima, proširuje se interesovanje korisnika ka susednim oblastima.

Prvih n preporuka.

- Korisnici uzimaju u obzir samo nekoliko sugestija, odnosno preporuka.
- Shodno tome, potrebno je što bolje rangirati preporuke.



Kako radimo?

- Način rada zavisi od domena i konkretnog problema.
- Trenutno, najbolji pristup je kolaborativno filtriranje.
- Drugi pristupi (preporuke zasnovane na sadržaju, kontekstno-zavisne preporuke, itd.) se mogu kombinovati da bi se poboljšali rezultati.
- U nastavku izlaganja obradićemo:
 - kolaborativno filtriranje i
 - preporuke zasnovane na sadržaju.

Šta je kolaborativno filtriranje?

- Kolaborativno filtriranje (engl. *collaborative filtering*, CF) je zadatak predviđanja (filtriranja) značaja novih artikala za korisnika prikupljanjem informacija o ukusu od velikog broja korisnika (kolaborativno).
- Izazovi u kolaborativnom filtriranju su:
 - veliki broj stavki (artikala) koje možete odabrati,
 - mali broj preporuka za predlaganje,
 - malo podataka po korisniku,
 - nepostojanje podataka za novog korisnika, i
 - veoma veliki skup podataka.

Tipovi kolaborativnog filtriranja.

- Postoje sledeći tipovi kolaborativnog filtriranja:
 - memorijski-zasnovano kolaborativno filtriranje (korisnički-zasnovano kolaborativno filtriranje, artikal-zasnovano kolaborativno filtriranje),
 - kolaborativno filtriranje zasnovano na modelu.

Korisnički zasnovano kolaborativno filtriranje.

- K1, K2, ..., K6 su korisnici.
- A1, A2, ..., A6 su artikli.
- Svaki korisnik je izrazio mišljenje za neke artikle.
- EksPLICITNO mišljenje se odnosi na ocenu artikla.
- IMPLICITNO mišljenje se odnosi na činjenicu da li je korisnik konzumirao artikl (na primer, pogledao film, odslušao pesmu, itd.)

	A1	A2	A3	A4	A5	A6
K1	2			4	5	
K2	5		4			1
K3			5		2	
K4		1		5		4
K5			4			2
K6	4	5		1		

Korisnički zasnovano kolaborativno filtriranje.

- Ciljni, odnosno aktivni korisnik je korisnik za koga obavljamo zadatak CF preporuke.

	A1	A2	A3	A4	A5	A6
K1	2			4	5	
K2	5		4			1
K3			5		2	
K4		1		5		4
→ K5			4			2
K6	4	5		1		

Korisnički zasnovano kolaborativno filtriranje.

- 1. Identifikujemo skup artikala koje je ciljni korisnik ocenio.

	A1	A2	A3	A4	A5	A6
K1	2			4	5	
K2	5		4			1
K3			5		2	
K4		1		5		4
→ K5			4			2
K6	4	5		1		

Korisnički zasnovano kolaborativno filtriranje.

- 1. Identifikujemo skup artikala koje je ciljni korisnik ocenio.
- 2. Identifikujemo susede – korisnike koji su ocenili jedan ili više artikala koji pripadaju tom skupu.

	A1	A2	A3	A4	A5	A6
K1	2			4	5	
K2	5		4			1
K3			5		2	
K4		1		5		4
→ K5			4			2
K6	4	5		1		

Korisnički zasnovano kolaborativno filtriranje.

- 3. Računamo funkciju sličnosti (engl. *similarity function*) između svakog suseda i ciljnog korisnika.
- 4. Po potrebi, bira se k najbližijih suseda.
- 5. Predviđaju se ocene artikala koje ciljni korisnik nije ocenio funkcijom predviđanja (engl. *prediction function*).
- 6. Ciljnom korisniku se preporučuju prvih (odnosno, po proračunu, za njega najznačajnijih) N proizvoda na osnovu proračunatih ocena.

Korisnički zasnovano kolaborativno filtriranje.

- Neka je u ciljni korisnik, Y matrica ocena, a $y_{v,i}$ ocena kojom je korisnik v ocenio artikal i .
- Funkcija sličnosti $sim(u, v)$ između korisnika u i v se računa na osnovu Pirsonove korelacije:

$$sim(u, v) = \frac{\sum_{i \in I_{uv}} (y_{u,i} - \hat{y}_u)(y_{v,i} - \hat{y}_v)}{\sqrt{\sum_{i \in I_{uv}} (y_{u,i} - \hat{y}_u)^2 \sum_{i \in I_{uv}} (y_{v,i} - \hat{y}_v)^2}}$$

- Funkcija predviđanja $y^*(u, i)$ data je jednačinom:

$$y^*(u, i) = \hat{y}_u + \frac{\sum_{j \in I_{y_j^* \neq 0}} sim(v_j, u) (y_{v_j, i} - \hat{y}_{v_j})}{\sum_{j \in I_{y_j^* \neq 0}} |sim(v_j, u)|}$$

Korisnički zasnovano kolaborativno filtriranje.

- Korisnike K1 i K6 ne uzimamo u obzir jer nisu susedi.

	A1	A2	A3	A4	A5	A6	$sim(u, v)$
K1	2			4	5		N/A
K2	5		4			1	
K3			5		2		
K4		1		5		4	
→ K5			4			2	
K6	4	5		1			N/A

Korisnički zasnovano kolaborativno filtriranje.

	A1	A2	A3	A4	A5	A6	$sim(u, v)$
K1	2			4	5		N/A
K2	5		4			1	0,87
K3			5		2		
K4		1		5		4	
→ K5			4			2	
K6	4	5		1			N/A

Korisnički zasnovano kolaborativno filtriranje.

	A1	A2	A3	A4	A5	A6	$sim(u, v)$
K1	2			4	5		N/A
K2	5		4			1	0,87
K3			5		2		1
K4		1		5		4	
→ K5			4			2	
K6	4	5		1			N/A

Korisnički zasnovano kolaborativno filtriranje.

	A1	A2	A3	A4	A5	A6	$sim(u, v)$
K1	2			4	5		N/A
K2	5		4			1	0,87
K3			5		2		1
K4		1		5		4	-1
→ K5			4			2	
K6	4	5		1			N/A

Korisnički zasnovano kolaborativno filtriranje.

	A1	A2	A3	A4	A5	A6	$sim(u, v)$
K1	2			4	5		N/A
K2	5		4			1	0,87
K3			5		2		1
K4		1		5		4	-1
→ K5	3,51*	3,81*	4	2,42*	2,48*	2	
K6	4	5		1			N/A

Artikal zasnovano kolaborativno filtriranje.

- U ovom slučaju, ciljni artikal je artikal za koji se obavlja CF predviđanje.
- Osnovni koraci su:
 - identifikovanje skupa korisnika koji su ocenili ciljni artikal,
 - identifikovanje drugih artikala (suseda) koje su ocenili korisnici iz prethodnog skupa,
 - računanje sličnosti između svakog suseda i ciljnog artikla (funkcija sličnosti),
 - po potrebi, bирамо k najbližnjih suseda, i
 - predviđamo ocene za ciljni artikal funkcijom predviđanja.

Artikal zasnovano kolaborativno filtriranje.

- Neka je i ciljni artikal, $y_{u,j}$ ocena kojom je korisnik u ocenio artikal j , a $\hat{y}_j$ prosečna ocena za j .
- Funkcija sličnosti $sim(i, j)$ između artikla i i j se računa na osnovu Pirsonove korelacije:

$$sim(i, j) = \frac{\sum_{u \in I_{ij}} (y_{u,i} - \hat{y}_i)(y_{u,j} - \hat{y}_j)}{\sqrt{\sum_{u \in I_{ij}} (y_{u,i} - \hat{y}_i)^2 \sum_{v \in I_{ij}} (y_{v,j} - \hat{y}_j)^2}}$$

- Funkcija predviđanja ocene $y^*(u, i)$ data je jednačinom:

$$y^*(u, i) = \hat{y}_i + \frac{\sum_{v \in I_{y_u^* \neq 0}} sim(i, j_v)(y_{u,j_v} - \hat{y}_{j_v})}{\sum_{v \in I_{y_u^* \neq 0}} |sim(i, j_v)|}$$

Artikal zasnovano kolaborativno filtriranje.

↓

	A1	A2	A3	A4	A5	A6
K1	2			4	5	
K2	5		4			1
K3			5		2	
K4		1		5		4
K5			4			2
K6	4	5		1		

$sim(u, v)$

Artikal zasnovano kolaborativno filtriranje.

↓

	A1	A2	A3	A4	A5	A6
K1	2			4	5	
K2	5		4			1
K3			5		2	
K4		1		5		4
K5			4			2
K6	4	5		1		

$sim(u, v)$ N/A

$sim(u, v)$	-1	N/A
-------------	----	-----

Artikal zasnovano kolaborativno filtriranje.

↓

	A1	A2	A3	A4	A5	A6
K1	2			4	5	
K2	5		4			1
K3			5		2	
K4		1		5		4
K5			4			2
K6	4	5		1		
$sim(u, v)$	-1	-1			N/A	

Artikal zasnovano kolaborativno filtriranje.

↓

	A1	A2	A3	A4	A5	A6
K1	2			4	5	
K2	5		4			1
K3			5		2	
K4		1		5		4
K5			4			2
K6	4	5		1		
$sim(u, v)$	-1	-1	0,86		N/A	

Artikal zasnovano kolaborativno filtriranje.

↓

	A1	A2	A3	A4	A5	A6
K1	2			4	5	
K2	5		4			1
K3			5		2	
K4		1		5		4
K5			4			2
K6	4	5		1		
$sim(u, v)$	-1	-1	0,86	1	N/A	

Artikal zasnovano kolaborativno filtriranje.

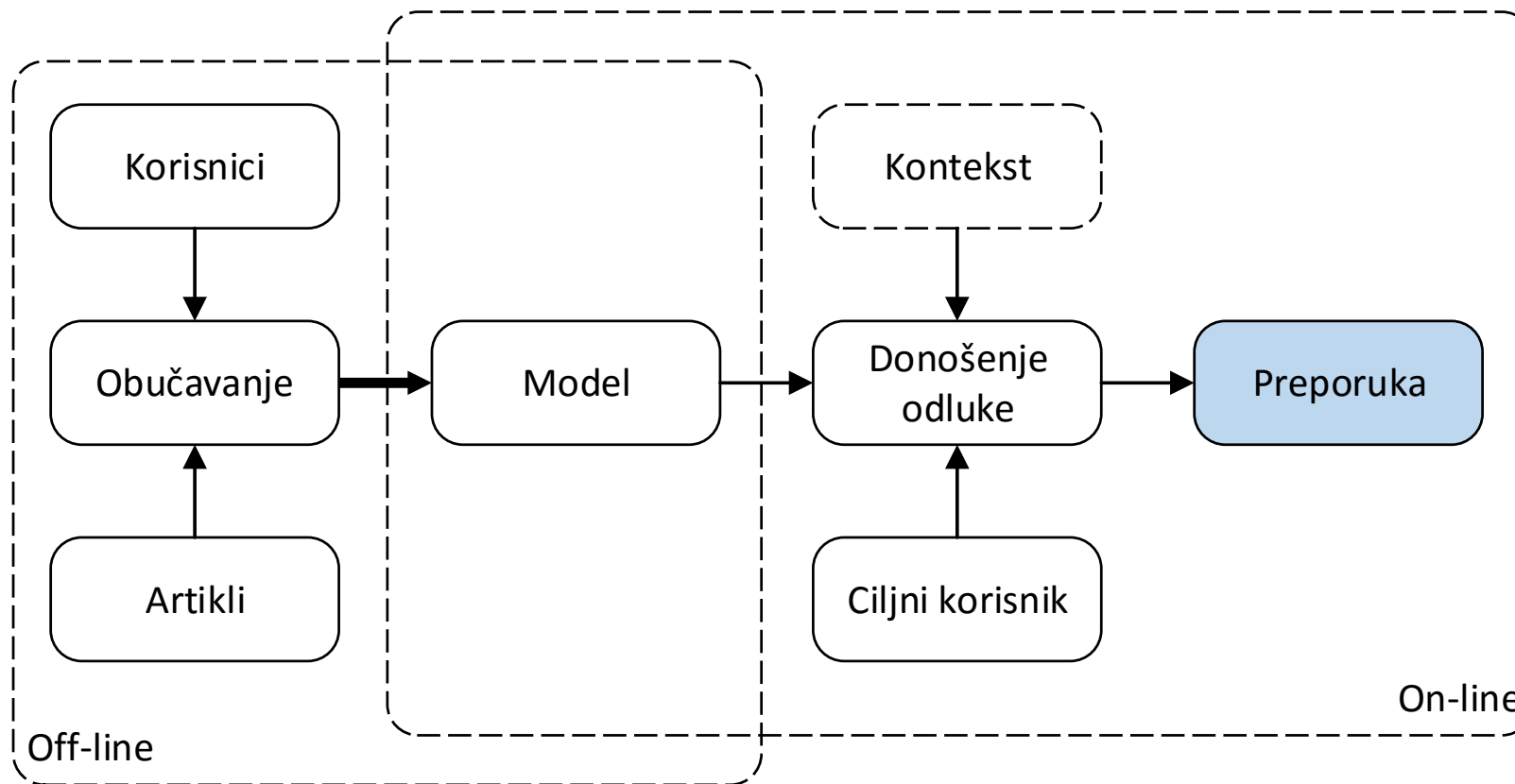
↓

	A1	A2	A3	A4	A5	A6
K1	2			4	5	2,94*
K2	5		4			1
K3			5		2	2,48*
K4		1		5		4
K5			4			2
K6	4	5		1		1,12*
$sim(u, v)$	-1	-1	0,86	1	N/A	

Uticaj na performanse.

- Kolaborativno filtriranje se odvija u dve faze:
 - u prvoj, koja se obavlja *off-line*, se računaju sličnosti, koje se potom čuvaju,
 - u drugoj, koja se obavlja *on-line*, se obavlja predviđanje.
- Usko grlo kolaborativnog filtriranja je računanje sličnosti.
- Proces je vremenski zahtevan zato što se izračunavanje vrši na osnovu podataka o velikom broju korisnika i velikom broju artikala.

Dve faze kolaborativnog filtriranja.



Uticaj na performanse.

- Određivanje suseda unapred može dovesti do lošeg predviđanja, pa se korisnički zasnovano računanje sličnosti obavlja više dinamički.
- Artikel zasnovano računanje sličnosti je više statičko, jer je moguće je odrediti susedne artikle unapred (samo se proces predviđanja obavlja *on-line*).

Memorijski zasnovano kolaborativno filtriranje.

- Dobre osobine su:
 - zahteva minimum inženjerskog znanja i truda,
 - korisnici i artikli su simboli bez ikakve unutrašnje strukture ili karakteristika,
 - proizvodi dobre rezultate u većini slučajeva.
- Loše osobine su:
 - potreban je veliki broj pouzdanih eksplicitnih ocena,
 - potrebno je da proizvodi budu standardizovani (npr., korisnici bi morali da kupe isti proizvod),
 - pretpostavlja da je trenutno ponašanje određeno prethodnim bez uzimanja u obzir znanja o kontekstu.

Problem proređenosti.

- Problem proređenosti, tj. raštrkanosti (engl. *sparsity problem*) se odnosi na velike skupove artikala (proizvoda) i mali broj korisničkih ocena.
- Primer 1.
 - U katalogu se nalazi milion knjiga.
 - Dva kupca su kupili po 100 knjiga.
 - Verovatnoća da su kupili istu knjigu je 0,01.
- Primer 2.
 - U katalogu se nalazi 10 miliona knjiga.
 - Dva kupca su kupili po 50 knjiga.
 - Verovatnoća da su kupili istu knjigu je 0,0002.
- Za kolaborativno filtriranje je potrebno da broj korisnika bude reda veličine 10% broja artikala.
- Problem je rešiv metodama smanjivanja dimenzionalnosti, klasterovanjem, itd.

Kolaborativno filtriranje zasnovano na modelu.

- Modeli predviđanja se formiraju na osnovu podataka i algoritama za mašinsko učenje i rudarenje podataka, a ne na osnovu jednostavne mere sličnosti.
- Za kolaborativno filtriranje zasnovano na modelu koristi se:
 - klasterovanje,
 - pravila udruživanja (engl. *association rules*),
 - ograničene Bolcmanove mašine (engl. *Restricted Boltzman Machines*, RBM),
 - rekurentne neuronske mreže,
 - Bajesove mreže, itd.
- U ovom izlaganju obratićemo pažnju na klaster-pristup, pravila udruživanja i RBM.

Klasterovanje.

- Kupci se grupišu u klastere na osnovu sklonosti i artikala koje su ranije kupili.
- Preporuke se izračunavaju na nivou klastera – svi kupci koji pripadaju jednom klasteru dobiće iste preporuke.
- Dobra osobina klasterovanja je što je računanje brzo, jer se izvodi na nivou klastera.
- Loše je to što su preporuke manje personalizovane.

Klasterovanje.

	Knjiga 1	Knjiga 2	Knjiga 3	Knjiga 4	Knjiga 5	Knjiga 6
Kupac A	X			X		
Kupac B		X	X		X	
Kupac C		X	X			
Kupac D		X				X
Kupac E	X				X	

- Preporuke za klaster kome pripadaju kupci 2, 3 i 4 bi mogle biti:
 - knjiga 2 (vrlo visoka),
 - knjiga 3 (visoka),
 - knjige 5 i 6 (mogu biti preporučene),
 - knjige 1 i 4 (nisu preporučene).

Pravila udruživanja.

- Pravila udruživanja (engl. *association rules*) su izjave oblika *ako-tada* (engl. *if-then*) koje pomažu u otkrivanju odnosa između naizgled nepovezanih podataka u nekom repozitorijumu.
- Primeri pravila udruživanja su:
 - „ako neko kupi deset jaja, on će sa verovatnoćom 80% kupiti i mleko“,
 - „ako neko kupi četiri piva, on će sa verovatnoćom 70% kupiti i čips“, itd.
- Pravila udruživanja su korisna za analizu i predviđanje ponašanja kupaca.
- Ona imaju važnu ulogu u analizi podataka o kupovini, grupisanju proizvoda, dizajniranju kataloga, itd.

Pravila udruživanja.

- Pravilo udruživanja ima dva dela, uzročni (ako) i posledični (tada).
 - Uzročni deo je stavka koja se nalazi u podacima.
 - Posledica je stavka koja je u vezi sa uzročnim delom.
- Pravila udruživanja generišu se analizom podataka (pronalaženjem učestalih obrazaca tipa ako-onda) i koristeći kriterijume podrška i poverenje za identifikaciju najvažnijih zavisnosti.
- Podrška (engl. *support*) je indikacija koliko često se stavke pojavljuju u bazi podataka.
- Poverenje (engl. *confidence*) se odnosi na broj puta kada su ako-onda izjave bile tačne.

Pravila udruživanja.

- Formalno: pravilo $X \rightarrow Y$.

$$supp = \frac{freq(X, Y)}{n}$$

$$conf = \frac{freq(X, Y)}{freq(X)}$$

$$lift = \frac{supp}{supp(X) \times supp(Y)}$$

Pravila udruživanja.

- Generisanje ako-onda pravila izvršava se u dva koraka:
 - pronalaženje onih skupova stavki čija verovatnoća dešavanja dostiže predhodno definisani prag podrške (ovakvi skupovi nazivaju se frekventni skupovi stavki), i
 - generisanje pravila udruživanja iz predhodno dobijenih velikih skupova stavki, a koja su ograničena sa minimalnom podrškom.

Pravila udruživanja.

- Apriori algoritam iterativno smanjuje minimalnu podršku dok ne pronađe pravila koja imaju poverenje jednako ili veće od postavljenog minimalnog poverenja.
- Prediktivni Apriori algoritam generiše n pravila za koje može biti postignuta maksimalna tačnost predviđanja. Neophodno je unapred definisati broj pravila (n), ali nije potrebno postavljati vrednosti za minimalnu podršku i poverenje kao u slučaju Apriori algoritma.
- Značaj prediktivnog Apriori algoritma je u činjenici da je u pitanju dinamička tehnika određivanja pravila koja koriste gornje granice tačnosti svih pravila.

Pravila udruživanja.

- Primer – uzročni deo.

	Knjiga 1	Knjiga 2	Knjiga 3	Knjiga 4	Knjiga 5	Knjiga 6
Kupac A	X			X		
Kupac B		X	X		X	
Kupac C		X	X			
Kupac D		X				X
Kupac E	X				X	
Kupac F			X		X	

Pravila udruživanja.

- Primer – posledični deo.

	Knjiga 1	Knjiga 2	Knjiga 3	Knjiga 4	Knjiga 5	Knjiga 6
Knjiga 1				1	1	
Knjiga 2			2		1	1
Knjiga 3		2			2	
Knjiga 4	1					
Knjiga 5	1		2			
Knjiga 6		1				

Pravila udruživanja.

- Dobre osobine pravila udruživanja su:
 - brzo se implementiraju,
 - brzo se izvršavaju,
 - ne zahtevaju mnogo prostora za skladištenje,
 - veoma su uspešna primenama koje se odnose na velike populacije, kao što je određivanje rasporeda polica u maloprodajnim objektima.
- Loše osobine pravila udruživanja su:
 - nisu pogodne za upotrebu ukoliko se sklonosti kupaca često menjaju,
 - mogu se koristiti samo u slučajevima da ima dovoljno podataka kojima se pravila mogu validirati (u suprotnom se mogu generisati lažne asocijacije).

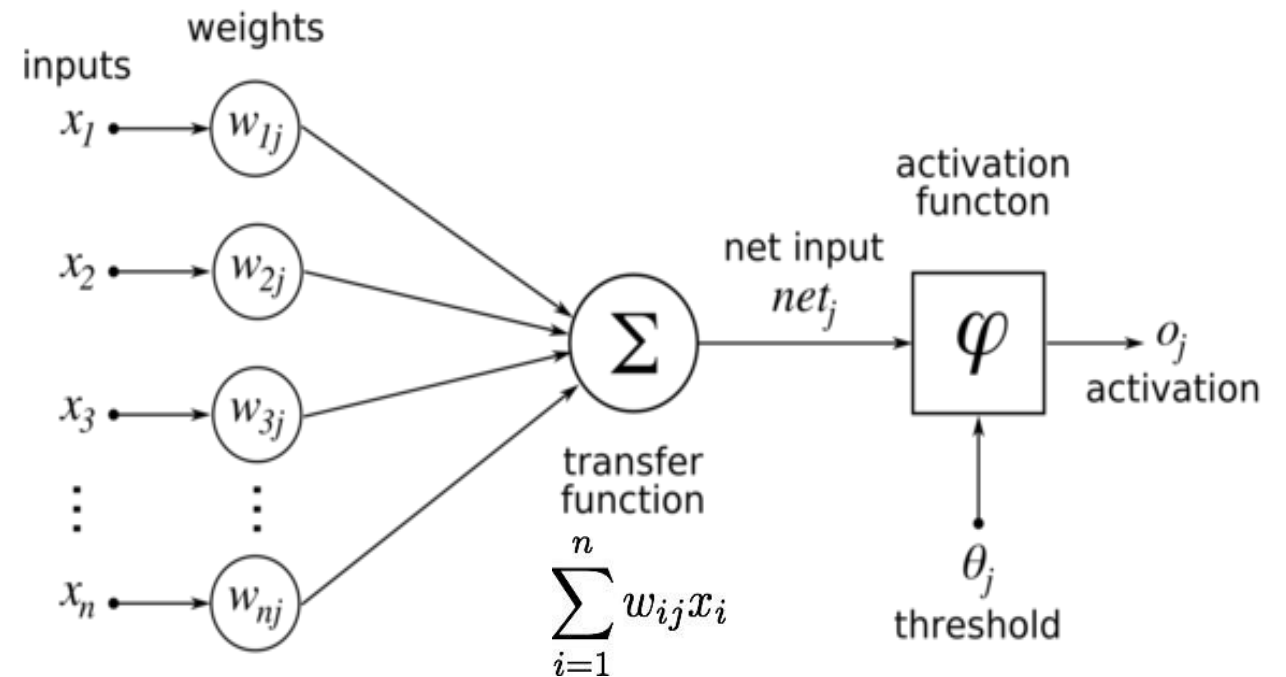
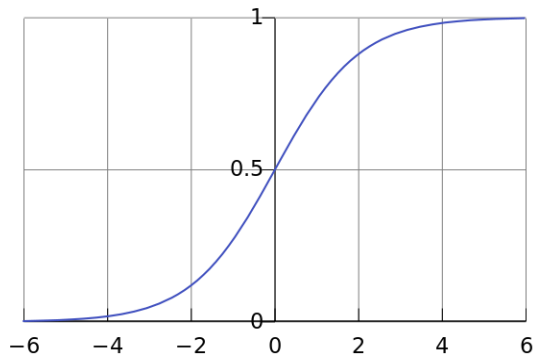
RBM.

- Ograničena Bolcmanova mašina (engl. *Restricted Boltzman Machine*) je (generativna stohastička) neuronska mreža koja uči raspodelu verovatnoće preko svojih ulaza.
- Koriste se za modeliranje tema, davanje preporuka itd.
- Osnovne su komponente nekih metoda dubokog učenja.

RBM.

$$P(o_j = 1|x) = \phi\left(\sum_{i=1}^n w_{ij}x_i + \theta_j\right)$$

$$\phi = \frac{1}{1 + e^{-\Sigma}}$$

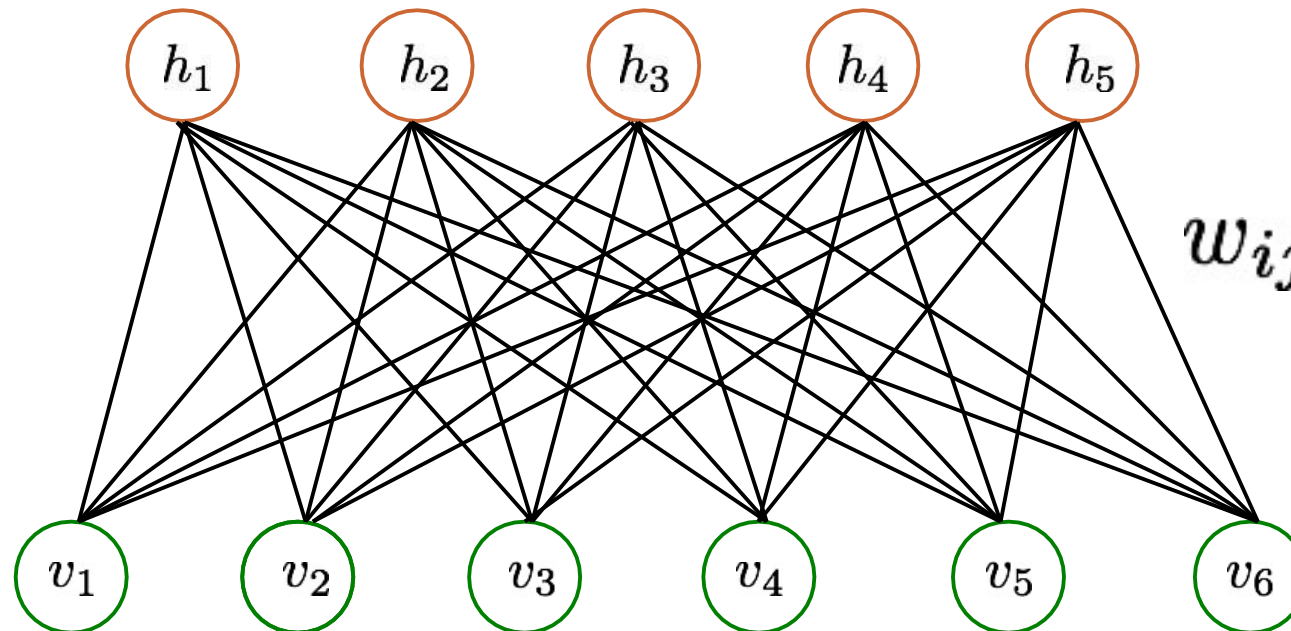


RBM.

- Svaka jedinica (engl. *unit*) je u stanju koje može biti aktivno ili neaktivno.
- Svaki ulaz u jedinicu je povezan sa težinskim koeficijentom (engl. *weight*).
- Funkcija prenosa (engl. *transfer function*) računa za svaku jedinicu rezultat kao sumu proizvoda težinskih koeficijenata i ulaza.
- Rezultat se prosleđuje aktivacionoj funkciji koja određuje verovatnoću da se jedinica nalazi u aktivnom stanju.

RBM.

- Svaka jedinica u vidljivom sloju v_i odnosi se na jedan artikal (proizvod).
- Svaka v_i povezana je sa skrivenom jedinicom h_i težinskim koeficijentom w_{ij} , pri čemu je broj skrivenih jedinica parametar.



RBM.

- U fazi obuke, za svakog korisnika se radi sledeće:
 - ukoliko je korisnik kupio artikl, aktivira se odgovarajuća v_i ;
 - aktivaciona stanja svih v_i su ulazi za svaku h_j ;
 - na osnovu ovih ulaza računa se aktivaciono stanje svake h_j ;
 - aktivaciona stanja svih h_j postaju novi ulazi svake v_i ;
 - aktivaciona stanja svake v_i se ponovo računaju;
 - za svaku v_i se na osnovu razlike između trenutnog i prethodnog aktivacionog stanja uređuju vrednosti težinskih koeficijenata w_{ij} i praga θ_j .

RBM.

- U fazi predviđanja, obučena RBM prilikom generisanja preporuke radi sledeće:
 - za sve article koji odgovaraju korisniku aktivira se odgovarajuća v_i ;
 - aktivaciona stanja svih v_i su ulazi za svaku h_j ;
 - na osnovu ovih ulaza računa se aktivaciono stanje svake h_j ;
 - aktivaciona stanja svih h_j postaju novi ulazi svake v_i ;
 - aktivaciona stanja svake v_i se ponovo računaju;
 - verovatnoće aktivacije se koriste za preporučivanje artikala.

Ograničenja kolaborativnog filtriranja.

- Zahteva podatke koji na neki način vezuju korisnike i artikle.
 - Potrebno je da u sistemu bude dovoljno korisnika.
 - Novi artikli trebaju da budu dovoljno ocenjeni, inače sistem neće moći da ih preporuči.
 - Novi korisnici moraju da ocene dovoljan broj artikala.
 - Problem sa novim artiklima i korisnicima je tzv. problem hladnog starta (engl. cold start).
- Raštrkanost – teško je pronaći korisnike koji su ocenili iste artikle.
- „Pristrasnost ka popularnosti“ – ne mogu se dati kvalitetne prepreke korisnicima sa vrlo specifičnim ukusom jer kolaborativno filtriranje favorizuje popularne artikle.

Šta su sistemi za preporuku zasnovani na sadržaju?

- Preporuke se zasnivaju na informacijama o sadržaju artikla (stavki), a ne na stavovima, odnosno ocenama drugih korisnika.
- Ovi sistemi koriste algoritam za mašinsko učenje kako bi izgradili model specifičan za datog korisnika na osnovu primera zasnovanih na opisu sadržaja artikala.

Šta je sadržaj artikla?

- Eksplicitni atributi ili karakteristike – na primer, za film:
 - žanr – akcija/avantura,
 - glavna uloga – Brus Vilis,
 - godina – 1995.
- Tekstualni sadržaj – na primer, za knjigu: naslov, opis, sadržaj.

Preporuke zasnovane na sadržaju

Profil.

- Preporučeni artikli (stavke) zasnovani su na profilu koji je napravljen analizom sadržaja stavki koje su se korisniku dopale u prošlosti.



Za šta je ovakav sistem pogodan?

- Ovakav sistem je pogodan za artikle, odnosno stavke koje su dobrim delom zasnovane na tekstu (knjige, Veb stranice, itd.)
- Artikli su opisani svojim obeležjima (npr. ključnim rečima).
- Korisnici su takođe opisani ključnim rečima u okviru artikala koje su kupili.
- Preporuke se daju na osnovu poređenja sadržaja (ključnih reči) artikala i korisnika.
- Model kojim je opisan korisnik takođe može biti i klasifikator (neuronska mreža, metoda vektora oslonca, naivni Bajes, itd.)

Dobre i loše osobine pristupa davanja preporuka zasnovanog na sadržaju.

- Dobre osobine:
 - Nisu potrebni podaci o drugim korisnicima.
 - Ne postoje problemi hladnog starta i raštrkanosti.
 - Mogu se preporučiti novi i nepopularni artikli.
 - Mogu se dati prepreke korisnicima sa vrlo specifičnim ukusom.
 - Korisnicima se može dati objašnjenje zašto su te preporuke date, npr. izlistavanjem ključnih reči.
- Loše osobine:
 - Neke vrste artikala (na primer, filmovi, muzika, itd.) nisu podložne jednostavnim metodama ekstrakcije obeležja.
 - Ukusi korisnika moraju se predstaviti kao funkcija koja se može naučiti od sadržaja.
 - Teško je iskoristiti kvalitetne procene drugih korisnika.

Metode davanja preporuka zasnovane na sadržaju.

- Neka je $Content(s)$ profil artikla, tj. skup obeležja – ključnih reči – koje karakterišu artikal s .
- Težinski faktor w_{ij} je mera značaja reči k_j u dokumentu d_j .
- Neka je $ContentBasedProfile(c)$ profil korisnika c formiran na osnovu analize sadržaja prethodnih artikala upotrebom tehnika za analizu ključnih reči.
- Ne primer, $ContentBasedProfile(c) = (wc_1, wc_1, \dots, wck)$, je vektor težinskih koeficijenata u kome je wc_i značaj reči k_i korisniku c .
- Funkcija korisnosti $u(c, s)$ definisana je na sledeći način:

$$e(c, s) = score(ContentBasedProfile(c), Content(s))$$

- gde su $ContentBasedProfile(c)$ korisnika c i $Content(s)$ artikla s predstavljeni kao TF-IDF vektori težinskih faktora ključnih reči (engl. *keyword weights*).

Metode davanja preporuka zasnovane na sadržaju.

- Funkcija korisnosti $u(c, s)$ je obično predstavljena nekim heurističkim bodovanjem definisanim na nivou vektora, kao što je kosinusna mera sličnosti:

$$u(c, s) = \cos(w_c, w_s) = \frac{w_c \times w_s}{\|w_c\| \|w_s\|} = \frac{\sum_{i=1}^K w_{ic} w_{is}}{\sum_{i=1}^K w_{ic}^2 \sum_{i=1}^K w_{is}^2}$$

- Sistemi za preporuku su jedno važno polje primene mašinskog učenja.
- Ovi sistemi imaju potencijal da postanu značajni kao što je sada značajna pretraga.
- Iako su sistemi za preporuku zasnovani na tehnologijama koje su se dobro pokazale (kolaborativno filtriranje, preporuke zasnovane na sadržaju, itd.), još uvek postoje razna otvorena pitanja.

- Beleške pripremljene prema predavanju – Alexandros Karatzoglou (2015): „Machine Learning for Recommender Systems“. Dostupno na adresi:
<http://rp-www.cs.usyd.edu.au/~mlss/content/mlss2015-karatzoglou.pdf>

Pitanja su dobrodošla.