

Predlaganje filmova korišćenjem content based i collaborative filtering sistema

Ilija Malinović; Marina Jerković

Fakultet tehničkih nauka

Apstrakt

U današnjem svetu korisnik se susreće sa širokim izborom, bilo da je u pitanju neka kupovina ili pak gledanje filma ili čitanje knjige - spektar mogućnosti je velik. Kako korisnik ne bi trošio svoje vreme tražeći idealan izbor, osmišljeni su sistemi za preporuku - sistemi koji korišćenjem različitih algoritama filtriraju podatke i nastoje da korisniku predlože proizvode koji će najbolje odgovarati njegovoj preferenci. Cilj ovakvih sistema jeste da se korisniku obezbedi što bolji doživljaj aplikacije, što dovodi do korisnikovog zadovoljstva i češće upotrebe aplikacije. U našem projektu biće korišten **content based i collaborative filtering algoritmi preporuke**. Za oba algoritma prikupljeni su podaci o filmovima kao i korisničkim ocenama nekih od filmova.

Content based filtering

Ovaj algoritam zasniva se na predlaganju filmova koji su po žanru slični onima koje je korisnik ranije ocenio pozitivno. Za potrebe ovog algoritma postoje 2 vrste podataka - **korisnički profil i profil filma**. Korisnički profil sastoji se od filmova koje je korisnik pozitivno ocenio, a profil filma koristi kao relevantne podatke žanrove u koje taj film spada. Ovaj način filtriranja računa sličnost filmova kao vrednost kosinusa ugla između dva vektora filma. Računanjem kosinusne sličnosti i rangiranjem, algoritam predlaže 10 filmova koji najbolje odgovaraju profilu korisnika.

$$\text{sim}(A, B) = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|}$$



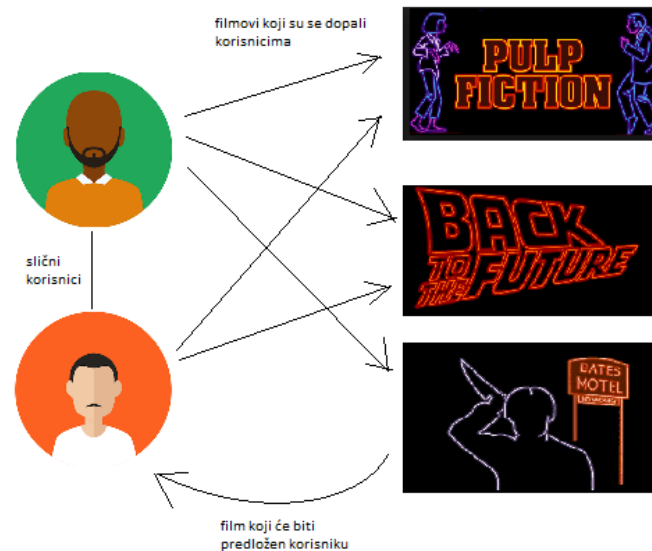
Collaborative filtering

Ovaj algoritam zasniva se na pronalaženju korisnika koji su slično ocenili filmove i predlaganju filmova koji su se svideli korisniku sa sličnim interesovanjima.

User-user filter traži sličnost među korisnicima na osnovu ocena koje su dali nekom filmu. Profil korisnika predstavlja vektor njegovih ocena za svaki film.

Item-item filter traži sličnost među filmovima na osnovu ocena koje su dobili od korisnika. Profil item-a predstavlja vektor dobijenih ocena od svakog korisnika.

I sličnost korisnika i sličnost filmova računaju se preko navedene kosinusne formule.



Evaluacija

Za evaluaciju collaborative filtera u projektu je korišten **RMSE** - root mean square error. Neke od ocena su izbrisane iz podataka, izračunata je matrica predikcije za postojeće podatke, pa su upoređene stvarne (izbrisane) ocene i predikovane ocene, izračunata je ukupna greška na osnovu formule sa slike.

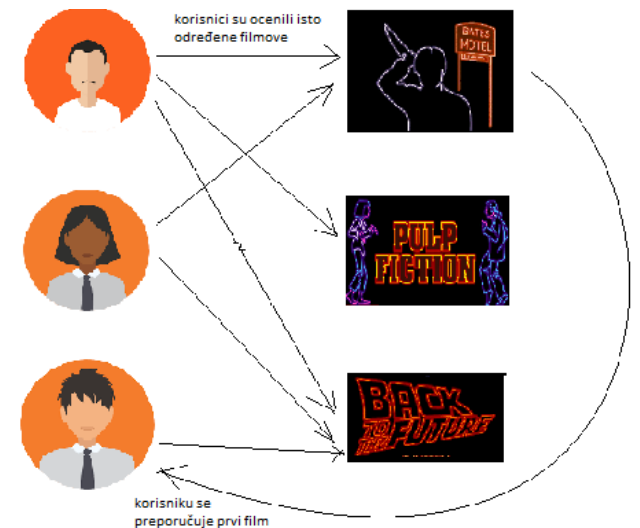
Računanje predikcije ocene vrši se preko formule na slici.

$$P_{u,i} = \frac{\sum_v (r_{v,i} * s_{u,v})}{\sum_v s_{u,v}}$$

$P_{u,i}$ predikcija ocene korisnika u za film i

$r_{v,i}$ rejting korisnika v za film i

$s_{u,v}$ sličnost korisnika u i v



$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2}$$

y_j predikovana ocena

\hat{y}_j stvarna ocena