

SISTEM ZA PREPORUKU FILMOVA

Projekat iz predmeta Osnovi računarske inteligencije

Asistent: Ivan Perić

Nikolina Radičić SW 27/2014 | Milan Šalić SW 53/2014

PODACI Obraden dataset preuzet sa MovieLens sajta. Csv fajl sadrži informacije o 9125 filmova. S obzirom da ti podaci nisu pružili dovoljno informacija, dodatne informacije o filmu su preuzete sa imdb profila filma. Drugi csv fajl sadrži 100004 ocjene od strane 671 razlicitog korisnika. Svi ponudjeni filmovi su bar jednom ocjenjeni. Odredjen broj filmova ne sadrzi podatke koji se za zadati film ne nalaze na imdb sajtu. Na osnovu navedenih informacija pravi se user-item matrica koja se popunjava informacijama učitanim iz fajla.

TRI METODE PREPORUKE

COLLABORATIVE FILTERING

Na osnovu podataka o tome kako su korisnici ocijenili filmove kreira se User-Item matrica X , gdje vrijednost $X[i,j]$ predstavlja kako je User i ocijenio film j . Na osnovu User-Item matrice pronalazimo slične korisnike po ocjenama i na osnovu njih predviđamo kako bi korisnik ocijenio ostale filmove. Ova metoda se može kombinovati sa preostale dvije, i tada njima prosljeđujemo 30 najboljih rezultata.

NEURAL NETWORK

Neuronsku mrežu obučavamo na 80% ocjena korisnika pri čemu za svaki ocijenjeni film imamo podatke o ocjeni, glumcima, režiseru, žanru, godini izdanja, bužetu, zaradi, zemlji porijekla i trajanju. Neuronska mreža vrši predikciju filmova klasifikacijom u grupe. Svaka grupa predstavlja određenu ocjenu od strane korisnika. Moguće ocjene filmova su: 0.5 , 1.0 , 1.5 , 2.0 , 2.5 , 3.0 , 3.5 , 4.0 , 4.5 , 5.0

REGRESION

Regresija takođe vrši predikciju ocjena od strane korisnika na osnovu informacija o filmovima prikupljenih sa Imdb sajta. Za razliku od neuronske mreže, predikcija od strane regresije ne vrši klasifikaciju po grupama. Predikcije koje dobijemo od regresije predstavljaju tačno predviđenu ocjenu u intervalu od 0-5, iako sistem za ocjenjivanje ne podržava tu mogućnost.

REZULTATI I ZAKLJUČAK

NAPOMENA: Za evaluaciju je izdvojeno 20% dataseta.

Rezultat Collaborative filtering

Collaborative filtering algoritam daje RMSE od 3.0311 na datasetu od 9125 filmova, 671 korisnikom i 100004 ocjene.

Rezultat Neural network

Neural network daje RMSE od 1.0630. Kada se kombinuje sa Collaborative filtering metodom p RMSE greška iznosi 1.0618

Rezultat Regresion

Regresion daje RMSE od 0.9560. Kada se kombinuje sa Collaborative filtering metodom RMSE greška iznosi 0,9534

Zaključak

Nemoguće je precizno odrediti tačnost rezultata svake metode zasebno , zbog toga što su metode međusobno povezane u jedan sistem. Sama tačnost rezultata nije toliko bitna, koliko je važan odnos ocjena između filmova. Na primjer, ukoliko program procjeni za sve filmove ocjenu više nego što je prava ocjena, naš sistem će i dalje ispravno raditi jer on bira najbolje filmove među svim filmovima i njih preporučuje korisniku. Mi mislimo da je to najbitnije prilikom rada programa.