

Sistem za preporuku filmova

Članovi tima: Luka Nikolić SW64/2017

Motivacija

Pronaći film koji nam se sviđa u današnje vreme sa izobiljem sadržaja je monumentalni zahvat, pogotovo za filmofile koji su već pogledali sve poznatije naslove. Potrebno je razviti aplikaciju koja će korisniku na osnovu njegovih preferencija i istorije pregleda predložiti najprikladnije filmove.

Opis problema

Problem predlaganja sadržaja su probale da reše mnoge kompanije i platforme. Netflix je razvio kompleksan machine learning algoritam sa velikim brojem parametara i podataka[1]. Koristeći Poasonovu korelaciju između filmova i regresione modele određuju jedinstvene predloge za svakog korisnika. Sa druge strane, rad [2] uvodi rule based model koji predviđa preferencije grupe korisnika na osnovu njihovih istorija i ocena.

Predlažem moj sistem za preporuku filmova koji će, koristeći Drools rule engine pored korisničkih omiljenih žanrova i ocena, uzeti u obzir ključne reči koje dublje opisuju filmove. Ključne reči su generalan pojam koji slobodnije i preciznije mogu da opišu atmosferu, iskustvo i osećanja dok gledamo film od drugih metrika. Iz tog razloga smatram da mogu da se koriste za kvalitetnije predviđanje filmova koji bi se korisniku svideli.

Metodologija rada

Za rešavanje ovog problema koristiće se sledećih entiteta:

- User: Predstavlja korisnika, sadrži lične podatke i bias podešavanje (HIGH, LOW), sviđanja, nesviđanja, pogledani filmovi i *watchlist*
- Movie: Predstavlja film iz baze koji se učitava na startupu. Sadrži informacije o režiseru, glumcima, trajanju, prosečnoj oceni.
- UserMovie: Predstavlja pogledan film od strane korisnika i sadrži weight(značaj) i da li je korisnik stavio da mu se sviđa ili ne sviđa film.
- Recenzije: Ostavljene od strane korisnika. Ocena utiče na weight odgovarajućeg filma
- Recommendation: Izlaz iz sistema, konkretni predlozi filmova sa bodovima
- Genre: Korisnik može da unese omiljene žanrove

Baza znanja:

Sistemu je potrebna kolekcija filmova koja se može prikupiti pomoću skrejpера sa online baza podataka filmova. Takođe su potrebne kolekcije ključnih reči i žanrova vezanih za konkretne filmove, koje se mogu prikupiti na isti način. Rule engine će učitati sve činjenice prilikom pokretanja aplikacije i formiranja baze. Nakon bilo koje izmene postojećih činjenica, baza podataka aplikacije će takođe biti ažurirana.

Forward Chaining:

Koristi se za preporuku na osnovu aktivnosti kao što su ostavljanje ocene ili dodavanje filma u pogledane

Preporuka filmova na osnovu pogledanih filmova(Pretpostavka da je sistem vec popunjen Movies cinjenicama):

1. Ako se registruje dodavanje filma u pogledane, modifikujemo korisnika.
2. Ako se registruje da je promenjena lista pogledanih filmova za korisnika, računaju se nove preporuke filmova(Recommendation)
 - a. Izbacujemo sve pogledane filmove iz razmatranja
 - b. Svaki predlog upoređujemo sa pogledanim filmovima i dodeljujemo bodove na osnovu sličnosti i težine koju smo dodelili svakom UserMovie
3. Ako se detektuju promene Recommendation cinjenica, sortiramo listu i prikazujemo prvih n

Backward Chaining:

Korisnik može da pretražuje filmove na osnovu konkretne pretrage sa raznim parametrima(žanr, godina, ocena, ...)

Template:

Za svakog korisnika se računa težina ili bias za svaki film koji je pogledao i to određuje koliki uticaj ima dati film na predloge novih filmova. Korisnik može da izabere High ili Low bias. Administrator može preko template-a da definiše High i Low profile i njihove konkretne vrednosti i uticaj na *weight pogledanog filma*.

CEP:

Primer 1:

Pratimo korisnike koji ostave više od 5 loših recenzija u kratkom periodu i onemogućimo mu ostavljanje recenzija

Primer 2:

Pratimo filmove koji imaju najviše pregleda i na početku svakog dana/meseca označimo kao *Popular* i bolje rangiramo u procenama

Konkretan primer

Ulazi korisnika

- Omiljeni žanrovi: Horror, Mystery, Drama
- Ključne reči: suspenseful, slasher, serial killer
- Nema recenzija i nema istoriju gledanja

Baza znanja sadrzi sledeće filmove: Silence of the Lambs (7.5), Inception(8.2), Airplane(7.9)

Izvršavanje pravila:

A1 - Filtriraju se filmovi iz baze znanja na osnovu izabranih žanrova. Rezultat ovog koraka su filmovi Silence of the Lambs i Inception

A2 – Na osnovu recenzija svih korisnika Inception dobija više poena (8.2)

A3 – Na osnovu ključnih reči suspenseful i serial killer Silence of the Lambs dobija 2 poena (ukupno 9.5), a Inception 1 zbog suspenseful (ukupno 9.2) Pošto korisnik nema istoriju, ne izvršava se korak A4 i najbolji film je Silence of the Lambs zbog najviše bodova

Literatura

1. Bennett, James, and Stan Lanning. "The netflix prize." Proceedings of KDD cup and workshop. Vol. 2007. 2007. Retrieved from: <https://www.cs.uic.edu/~liub/KDD-cup-2007/NetflixPrize-description.pdf>
2. Sharma, Mugdha, Laxmi Ahuja, and Vinay Kumar. "A novel rule based data mining approach towards movie recommender system." Journal of Information and Organizational Sciences 44.1 (2020): 157-170. Retrieved from: <https://hrcak.srce.hr/file/348076>