

Sistem za preporuku knjiga

Motivacija

Sistemi za preporuku su od velikog značaja u industriji.

Pomažu korisnicima da dođu do željenog sadržaja.

U pozadini ovih je funkcija preporuke. Ona uzima informaciju o korisniku i predviđa sadržaj koji bi mu se sviđao.

Podaci

Koristili smo "goodbooks-10k" skup podataka. On sadrži spisak od 10 000 najpopularnijih knjiga sa sajta Goodreads.com i 6 000 000 korisničkih ocena. Takođe, tu se nalazi i spisak korisnički definisanih etiketa koje opisuju knjige.

Podatke smo morali predprocesirati. Uklonili smo duplike ocene, uvezvi poslednju u obzir. Takođe, izbacili smo etikete koje nisu na engleskom jeziku kako bi modeli preporuke učinili preciznijim. Na kraju, spojili smo etikete vezane za određenu knjigu u poseban fajl.

Metode

Najpopularnije knjige

Dve metode preporuke koje nisu personalizovane.

Ne iziskuju nikakvo treniranje već samo podatke.

Poželjno ih je koristiti ako nemamo nikakve podatke o trenutnom korisniku.

Najbolje ocenjene knjige

Content based filtering je grupisanje knjiga na osnovu sličnosti njihovih etiketa. Preporučivanje na osnovu takve kategorizacije.

Collaborative filtering

Predstavlja skup postupaka koji za predlaganje knjiga koriste prethodno date ocene.

Na osnovu njih, gleda se sličnost između korisnika ili knjiga. Mi smo se odlučili za predlaganje na osnovu prvog.

Singular Value Decomposition (SVD)

Koristimo matematičku osobinu da se svaka matrica može faktorizovati na tri manje.

Matematički, ovo je izraženo formulom: $A = U \Sigma V^T$, gde je:

- A matrica mxn
- U ortogonalna mxm matrica
- Σ ne-negativna mxn dijagonalna matrica
- V ortogonalna mxn matrica.

Ovim dobijamo matrice manje dimenzionalnosti.

U projektu to primenjujemo nad ocenama knjiga date od strane korisnika.

Tražimo sličnost "ukusa" između korisnika, na osnovu kovarijanse matrica, predlažući nove knjige na osnovu dosadašnjih ocena.

Koristimo implementaciju algoritma u surprise modulu.

Neural Collaborative Filtering

Implementacija modela predloženog u radu "Neural Collaborative Filtering" - 2017 He et al. u Keras okruženju.

Kombinujemo korisnike i njihove ocene u matrice malih dimenzija, tzv. "embedding".

Nakon toga, korišćenjem "Generalized Matrix Factorization" algoritma spajamo matrice njihovim množenjem. Priměnom višešlojnog perceptrona takođe dobijamo "embedding" ali ga posmatramo kao vektor osobina i posledujemo ga narednim slovima.

Nakon toga, naša mreža kombinuje obe rezultata i daje predlog knjiga.

Rezultati

Za Singular Value Decomposition i Neural Collaborative Filtering postupke, dobili smo sledeće ocene:

Singular Value Decomposition:
 $MSA = 0.640$
 $RMSE = 0.830$

Neural Collaborative Filtering:
 $MSA = 0.651$
 $RMSE = 0.834$

Procenu rada metode content based filtering postupka nismo uspeli izračunati.

Jedan od vidova evaluacije u industriji je praćenje procenta odziva korisnika na predložene knjige.