

# Veštačka inteligencija

## Projekat 2

Sistemi za preporuku (Recommender sistemi)

Božidar Tošić 18016  
Andrija Tošić 18015  
Stefan Stojadinović 17975

# Sadržaj

<b>Sistemi za preporuku (Recommender sistemi)</b>	<b>1</b>
Sadržaj	2
Uvod	3
Aktuelni projekti i tehnike	3
Filtriranje zasnovano na sadržaju	3
Saradničko filtriranje	3
Hibridni sistemi	3
Formulacija problema	3
Kosinusna sličnost	4
Dekompozicija matrice $M$ na osnovu sopstvenih vrednosti	5
Kratak opis rešenja	5
Zahtevi kod pokretanja	5
Saradničko filtriranje	6
Koraci u implementaciji	6
Filtriranje zasnovano na sadržaju	6
Koraci u implementaciji	6
Reference	7

## Uvod

Sistemi preporuka su kompjuterski programi koji korisnicima predlažu preporuke u zavisnosti od niza kriterijuma. Ovakvi sistemi poboljšavaju korisničko iskustvo krajnjim korisnicima, a osim toga omogućavaju lakšu prodaju usluga i proizvoda pružanjem preporuka korisnicima za koje su verovatno zainteresovani.

Sistemi preporuka obično rade sa velikom količinom informacija, a sa ciljem filtriranja preporuka. Osnovu podataka osim informacija o predmetima preporuka predstavljaju i informacije o korisničkim preferencama koje korisnici direktno saopštavaju sistemu i informacije koje sistem zaključuje o korisniku.

Konkretan problem koji pokušavamo da rešimo je koje pesme preporučiti na osnovu njihovih atributa i ocena koje je dao slušaoc pesmama.

## Aktuelni projekti i tehnike

Israživanjem i razvojem sistema preporuka bave se velike kompanije, a osim njihovih rešenja dostupni su i brojni opensource projekti.

Osnovne tehnike na kojima se baziraju ovakvi sistemi:

### Filtriranje zasnovano na sadržaju

Sistemi zasnovani na sadržaju generišu preporuke na osnovu prethodnog ponašanja i preferenca korisnika. Ovaj tip sistema se obično koristi za preporuku stavki koje su slične stavkama s kojima je korisnik ranije interagovao. Nivo sličnosti između stavki se generalno utvrđuje na osnovu atributa stavki.

### Saradničko filtriranje

Ovaj tip sistema daje preporuke na osnovu ponašanja sličnih korisnika. Algoritmi za kolaborativno filtriranje rade tako što analiziraju prethodno ponašanje korisnika i identifikuju druge korisnike sa sličnim ponašanjem. Postoje dva glavna tipa saradničkog filtriranja: zasnovano na korisniku i zasnovano na stavkama. Logika iza kolaborativnog filtriranja je da ako korisnici A i B imaju sličan ukus u vezi jednog proizvoda, onda će A i B verovatno imati sličan ukus i što se tiče drugih proizvoda.

### Hibridni sistemi

Ovaj tip sistema kombinuje prednosti filtriranja zasnovanog na sadržaju i saradničkog filtriranja. Koriste prethodno ponašanje i preference korisnika kao i ponašanje sličnih korisnika da daju preporuke. Ovaj tip sistema obično je precizniji i efikasniji nego samostalno saradničko ili filtriranje zasnovano na sadržaju.

## Formulacija problema

Koristićemo primer recommender sistema za preporuku pesama korisnicima.

Sama formulacija problema obično uključuje korišćenje istorijskih podataka o interakcijama korisnika (slušaoca pesama) sa skupom stavki da bi se izgradio model koji može predvideti koje sa kojim stavkama će korisnik verovatno imati interakciju u budućnosti. Osnovni cilj je povećanje tačnosti ovih predviđanja, tako da sistem preporuka bude u mogućnosti da pruži personalizovane preporuke korisnicima koje će im najverovatnije biti od koristi.

Skup podataka u sistemu za preporuku pesama mogu biti ocena pesme, godina izdavanja, izdavač, žanr, itd. Dok informacije o korisniku (slušaocu) mogu biti izvedene iz prethodno slušanih pesama, odabranih žanrova, broj godina. U svakoj vrsti csv fajla se nalaze jedinstvena kombinacija korisnika (slušaoca), pesme, njegova ocena i atributi pesme.

podaci.csv

	Id_pesme ▾	Id_izvodjaca ▾	Zanr_pesme ▾	Id_slusaoca ▾	Trajanje_pesme ▾	Ocena_pesme ▾	Godina_izdavar ▾	Jezik ▾
	329	61	7	2106	6	2	2002	5
	33	26	4	1996	1	8	2001	6
	262	41	10	552	3	3	2022	1
	435	166	8	2659	6	1	2022	6
	295	266	9	2197	3	3	2013	4
	435	282	9	279	5	9	2021	5
	371	21	2	1264	2	7	2012	6

## Kosinusna sličnost

U analizi podataka, kosinusna sličnost je mera sličnosti između dva niza brojeva.

Kosinus dva vektora različita od nule može se izračunati korišćenjem formule Euklidskog skalarnog proizvoda:

$$\mathbf{A} \cdot \mathbf{B} = \|\mathbf{A}\| \|\mathbf{B}\| \cos \theta$$

Za dva n-dimenzionalna vektora atributa, A i B, kosinusna sličnost,  $\cos(\theta)$ , predstavljena je korišćenjem skalarnog proizvoda i intenziteta kao:

$$S_C(A, B) := \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|}$$

Korišćenjem ove formule mogu se naći slični vektori koji sadrže vrednosti atributa pesama.

## Dekompozicija matrice M na osnovu sopstvenih vrednosti

$$\begin{array}{ccccc}
 \begin{array}{|c|c|c|c|} \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline \end{array} & 
 \begin{array}{|c|c|c|c|} \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline \end{array} & 
 \begin{array}{|c|c|c|} \hline & & \\ \hline & & \\ \hline & & \\ \hline & & \\ \hline \end{array} & 
 \begin{array}{|c|c|c|} \hline & & \\ \hline & & \\ \hline & & \\ \hline & & \\ \hline \end{array} \\
 \mathbf{M} & = & \mathbf{U} & \mathbf{\Sigma} & \mathbf{V}^* \\
 m \times n & & m \times m & m \times n & n \times n
 \end{array}$$
  

$$\begin{array}{ccc}
 \begin{array}{|c|c|c|c|} \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline \end{array} & 
 \begin{array}{|c|c|c|c|} \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline \end{array} & 
 \begin{array}{|c|c|c|c|} \hline 1 & 0 & 0 & 0 \\ \hline 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 1 & 0 \\ \hline 0 & 0 & 0 & 1 \\ \hline \end{array} \\
 \mathbf{U} & \mathbf{U}^* & = \mathbf{I}_m
 \end{array}$$
  

$$\begin{array}{ccc}
 \begin{array}{|c|c|c|} \hline & & \\ \hline & & \\ \hline & & \\ \hline & & \\ \hline \end{array} & 
 \begin{array}{|c|c|c|} \hline & & \\ \hline & & \\ \hline & & \\ \hline & & \\ \hline \end{array} & 
 \begin{array}{|c|c|c|} \hline 1 & 0 & 0 \\ \hline 0 & 1 & 0 \\ \hline 0 & 0 & 1 \\ \hline \end{array} \\
 \mathbf{V} & \mathbf{V}^* & = \mathbf{I}_n
 \end{array}$$

U linearnoj algebri, dekompozicija sopstvenih vrednosti (SVD) je faktorizacija realne ili kompleksne matrice. Konkretno, dekompozicija sopstvenih vrednosti kompleksne matrice  $M$  dimenzija  $m \times n$  je faktorizacija oblika  $M = U \Sigma V^*$ , gde je  $U$  kompleksna unitarna matrica dimenzija  $m \times m$ , sigma je pravougaona dijagonalna matrica dimenzija  $m \times n$  sa nenegativnim realnim brojevima na dijagonali,  $V$  je  $n \times n$  kompleksna unitarna matrica, a  $V^*$  je konjugovana transpozicija od  $V$ .

	Harry Potter	The Triplets of Belleville	Shrek	The Dark Knight Rises	Memento	
	✓		✓	✓		.9
		✓			✓	-1
	✓	✓	✓			1
				✓	✓	1
						-.9
						-.2
						-1
						.9
						1
						.88
						-1.08
						0.9
						1.09
						-0.8
						-0.9
						1.0
						-1.0
						-1.0
						0.9
						.2
						-1
						0.38
						0.6
						1.2
						-0.7
						-1.18
						.1
						1
						-0.11
						-0.9
						-0.9
						1.0
						0.91

Primer matrice ocena za korisnike i filmove i njene faktorizacije

## Kratak opis rešenja

### Zahtevi kod pokretanja

Potrebni Python paketi: [pandas](#), [numpy](#), [scikit-learn](#)

Instalirati komandom `pip install <paket>`

## Saradničko filtriranje

### Koraci u implementaciji


1. Podaci se uvoze u pandas DataFrame iz priloženog csv fajla
2. Generiše se pivot tabela u kojoj su čitaoci indeks, kolona su pesme, a vrednosti su ocene
3. Ocene pesama iz pivot tabele se konvertuju u retko posednutu matricu
4. Matrica sa ocenama pesama se faktoriše
5. Predviđene ocene se dobijaju skalarnim proizvodom dobijenih matrica iz faktORIZACIJE
6. Vektor s predviđenim ocenama se konvertuje u pandas DataFrame i pogodan je za prikaz

## Filtriranje zasnovano na sadržaju

### Koraci u implementaciji

1. Podaci se uvoze u pandas DataFrame iz priloženog csv fajla
2. Vrednosti u kolonama koje su atributi pesme se normalizuju
3. Nad kolonama se vrši one-hot kodiranje
4. Izračunavanje sličnosti vektora sa id-jevima pesma u odnosu na sve ostale vektore
5. U koloni za predikciju se upisuju vrednosti koje su rezultat kosinusne sličnosti

## Reference

1.  How Recommender Systems Work (Netflix/Amazon)
2. Korišćena open source Python biblioteka - [GitHub - NicolasHug/Surprise: A Python scikit for building and analyzing recommender systems](#)
3. [Book-Crossing: User review ratings | Kaggle](#)
4. [MATRIX FACTORIZATION TECHNIQUES FOR RECOMMENDER SYSTEMS](#)
5. <https://towardsdatascience.com/introduction-to-recommender-systems-6c66cf15ada>
6. <https://developers.google.com/machine-learning/recommendation>
7. <https://www.nvidia.com/en-us/glossary/data-science/recommendation-system/>