

Sistem za preporuku animea

Božidar Arsić

Fakultet tehničkih nauka
Univerzitet u Novom Sadu
Trg Dositeja Obradovića 6
21000 Novi Sad

Abstract— Anime, japanska forma animacije, postaje sve popularnija širom sveta, privlačeći publiku svojom jedinstvenom estetikom, dubokim pričama i raznovrsnim žanrovima. Ova forma umetnosti se sada promovise na mnogim platformama i događajima širom sveta. Popularnost animea stvorila je potrebu za personaliziranim preporukama sadržaja, a razvijanje efikasnog sistema preporuke predstavlja značajan izazov u polju filtriranja sadržaja. Cilj ovog istraživanja je razvijanje hibridnog sistema preporuke animea koji kombinuje Collaborative i Content-based načine filtriranja. Koristeći skup podataka koji sadrži ocene korisnika i metapodatke animea, implementirani sistem za preporuku uzima u obzir korisnikove lične preference i karakteristike animea kako bi predložio personalizirane preporuke. Evaluacija hibridnog sistema preporuke pokazala je bolje rezultate u odnosu na pojedinačne pristupe preporuke, te predstavlja potencijalno rešenje za izazove personalizirane preporuke animea. Rezultati ovog istraživanja mogu biti primijenjeni u različitim područjima, poput edukacije, e-trgovine, turizma i zabave, gdje se traži personalizirana preporuka sadržaja. Ova studija ima praktičnu primjenu u razvijanju efikasnih sistema preporuke za personaliziranu uslugu korisnicima, kao i u doprinosu razumijevanju popularnosti animea i njegovog utjecaja na savremeno društvo.

Ključne reči—hibrid; anime; sistem preporuke; fuzzy logika

I. UVOD

Eksplzivni rast količine dostupnih digitalnih informacija i broja posetilaca interneta stvorili su potencijalni problem preopterećenosti informacijama koje ometaju brz pristup informacijama od interesa. Takav problem iziskuje potrebu za filtriranjem sadržaja i efikasnim dostavljanjem relevantnih informacija.

Sistemi za pronalaženje informacija kao što su Google, DevilFinder i Altavista su delimično rešili ovaj problem, ali su izostavili prioritizaciju i personalizaciju sadržaja. Upravo zato, sistemi za preporuku postali su ključni deo brojnih aplikacija koje nude personalizovan sadržaj i usluge. Iako primarno područja u kojima je postojanje sistemi za preporuku već neophodno jesu platforme za strimovanje i sajtovi koji vrše usluge emitovanja serija i filmova, oni primenu i veliki značaj nude i u drugim oblastima.

Oni rešavaju probleme preopterećenosti informacijama, filtriranjem njihovih vitalnih fragmenata prema željama korisnika, interesovanjima, ili uočenog ponašanja u vezi sa određenim predmetom.

Sistem za preporuku ima mogućnost da predvidi da li će određenom korisniku, stavka ili predmet odgovarati u zavisnosti od njegovog profila.

Oni takođe smanjuju transakcione troškove pronalaženja i izbora artikala za kupovinu na okruženjima za online shopping. Pokazalo se da su sistemi preporuka poboljšali proces donošenja odluka i kvalitet.

U svetu e-trgovine, sistemi za preporuku su povećali prihode, zbog činjenice da su oni efikasno sredstvo za plasman više proizvoda.

U svetu naćnog sadržaja ili biblioteka, sistemi za pretragu pružaju podršku korisnicima da prevaziđu pretragu u vidu obićnih kataloga.

U ovom radu će sistem za preporuku biti primenjen na animeima. Detaljniji opis podataka, izazova i rešenja izložen je u ostatku rada. Naredno poglavlje se bavi srođnim istraživanjima na ovu temu. U trećem poglavlju opisuju se metodologije koje su korišćene. Nakon toga, biće predstavljeni rezultati i diskusije. Na kraju će biti izveden zaključak ovog rada i moguća unapređenja.

II. PREGLED LITERATURE

U radu [1], Isinkaye je opisano teorijske osnove sistema za preporuku. U njemu se jasno naglašavaju predosti i mane svake od mogućih tehnika preporuke, kao i moguće načine da se oni prevaziđu. Ovaj rad je dobra ulazna taćka koja daje jasnije smernice kako pristupiti problemu preporuke.

U radu [2], Reddy je opisao kako se može primeniti *Content-base* filtering na osnovu žanrova filmova. Ovaj jednostavan primer implementacije logike proširuje teorijsko znanje prvog rada praktićnim primerom.

U radu [3], Walek je prikazao kompletan način funkcionisanja hibridnog sistema. Ovaj rad je pružio najveću inspiraciju i smernice kako implementirati hibridni sistem. Nastavak rada se u velikoj meri bazira na arhitekturi i tehnikama koje su navedene u njemu. Zanimljiv pristup implementacije hibridnog sistema je uvođenje Expert sloja koji koristeći *fuzzy* logiku rukuje rezultatima *Collaborative* i *Content-base* slojeva. U radu se prikazuje kako takav način implementacije rešenja postiže izuzetno veliku taćnost preporuke.

III. METODOLOGIJA

A. Definisanje problema

Anime	K1	K2	K3	K4
A	1	3	5	5
B	?	?	2	?
C	5	2	2	?
D	5	1	4	5
E	1	?	?	4

Tabela 1 - Ocene animea od strane korisnika

Kako bismo razumeli problem koji rešavamo, možemo posmatrati *Tabela 1*. Ona sadrži 5 animea (A, B, C, D, E) i četiri korisnika (K1, K2, K3, K4). Korisnik K1 je dodelio ocene animeima A, C, D, E redom 1, 5, 5, 1. Korisnik je dodelio najviše moguće ocene animeima C i D, čime je naznačio da mu se oni dopadaju. Obrnuto tome, naglasio da mu se ne sviđaju animei A i E. Korisnik K1 nema dodeljenu ocenu za anime B, na osnovu čega ćemo pretpostaviti da on taj anime nije ni gledao. Koristeći istu logiku *Tabela 1* prikazuje ocene i drugih korisnika sistema (K2, K3 i K4) za svaki od ponuđenih animea (A, B, C, D, E). Sistem treba da preporuči svakom korisniku anime koji bi se njemu najviše svideo i za to koristi istorijske podatke ocena.

On to postiže tako što vrši kalkulacije kakvu bi ocenu korisnik dodelio svakom od neocenjenih animea i vraća sortiranu listu animea sa najvećom procenjenom ocenom.

Anime	Akcija	Komedija
A	DA	NE
B	DA	NE
C	DA	DA
D	NE	DA
E	DA	NE

Tabela 2 – Sadržaj animea

Dodatne informacije o animeima mogu nam pomoći da poboljšamo procenu korisnikovih preferencija. Kao primer možemo ponovo koristiti korisnika K1 i već poznati podatak da voli anime C i D, a da ne voli A i E. Ako pogledamo *Tabela 2* možemo doći do zaključka da korisnik preferira komedije kao žanr.

Kako bismo došli do kvalitetnijih procena možemo koristiti *Collaborative* ili *Content-Base* filtering.

B. Collaborative filtering

Collaborative filtering je tehnika koja može da filtrira produkte koje se korisniku možda dopadaju oslanjajući se na ocene drugih sličnih korisnika. Funkcioniše po principu analiziranja veće grupe korisnika kako bi pronašli manju grupicu koja ima slična interesovanja kao željeni korisnik. Tada, koristeći detektovanu grupu posmatramo kako je ona u proseku ocenila anime za koji predviđamo ocenu. Ako taj postupak primenimo na svim neocenjenim animeima za željenog korisnika, možemo dobiti listu preporuke tako što ćemo sortirati listu neocenjenih animea prema prediktovanoj oceni [4].

Postoje različiti načini implementiranja *Collaborative filtering* tehnike kao što su *Memory Based* i *Model based*.

U trenutnom radu korišćen je *Model based* pristup uz korišćenje *SVD* modela nad *user-item* matricom. Opis ovog dela logike je prikazan u tački D ovog poglavlja (*SVD model*)

C. Content-base filtering

Content-based tehnika je domenski-orijentisan algoritam koji se oslanja na attribute artikla kako bismo došli do predikcije. U slučaju animea, korisniku će se preporučivati animei sličnijeg sadržaja kao što su već pozitivno ocenjeni animei. Postoje različiti algoritmi za kalkulisanje sličnosti između sadržaja nekog artikla kao što su *Term Frequency Inverse Document* (TF-IDF), *Decision Trees*, *Naïve Bayes Classifier*, *Neural Networks* [3].

Prednost *Content-Base* filteringa je što ne zahteva postojanje drugih korisnika kako bi prediktovao ocenu. Takođe ako se korisnikov profil promeni, *Content-based* će potencijalno moći da doradi svoje preporuke za veoma kratko vreme.

Glavni nedostatak ovog algoritma jeste što će se korisniku preporučivati samo artikli za koje je već pokazao interesovanje i skoro je nemoguće da dobije preporuke van tog opsega.

D. Singular Value Decomposition model

Singular Value Decomposition (SVD) model predstavlja *Model based* način primene *Collaborative filtering* tehnike. Kao što smo naglasili, u ovom radu se koristi *user-item* matrica kako bi se odradila predikcija ocene animea. Takva matrica je identičnog izgleda kao što je *Tabela 1*. Kolone matrice su korisnici, redovi predstavljaju anime, a u preseku su ocene dodeljene konkretnom animeu od strane korisnika.

Realan problem koji se mora prevazići je ograničen broj resura kojim rukujemo u vidu memorije i činjenice da *user-item* matrica može lako da zauzme skoro sve resurse usled njene velike dimenzionalnosti.

Rešenje se pronalazi u korišćenju *SVD* modela koji ima zadatak da, koristeći računске operacije, *user-item* matricu „razbije“ na tri manje podmatrice i to tako da proizvodom te tri novokreirane matrice, dobijemo početnu *user-item* matricu [3].

$$A = U W V^T$$

A – originalna matrica dimenzionalnosti $m \times n$

U – matrica korisničkih karakteristika $m \times r$

W – dijagonalna matrica težina $r \times r$

V – matrica karakteristika animea (u jednačini ona se dodatno transponuje) $r \times n$

E. Fuzzy logika

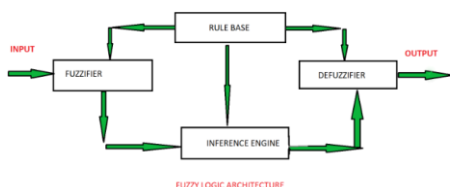
Fuzzy logika je oblik logike sa više vrednosti u kojoj istinite vrednosti promenljivih mogu biti bilo koji realni broj između 0 i 1, umesto samo tradicionalnih vrednosti „tačno“ ili „netačno“. Koristi se za rad sa nepreciznim ili nesigurnim

informacijama i matematički je metod za predstavljanje nejasnoće i neizvesnosti u donošenju odluka [5].

Fuzzy logika se implementira uvođenjem Fuzzy pravila koji predstavljaju izraze oblika *if-then* kao na primer „ako je plata mala, tada su šanse za odobrenje kredita mala“

Arhitekturu fuzzy modela (Slika 1) čine elementi:

- *Rule base* - Baza znanja koju čini skup if-then pravila
- *Fuzzification* – postupak konvertovanja grubih vrednosti brojeva u nejasne (*fuzzy*) vrednosti
- *Inference engine* – odlučuje o nivou poklapanja unete nejasne (*fuzzy*) vrednosti sa navedenim pravilima iz baze znanja
- *Defuzzification* – koristi se da konvertuje rezultujuće nejasne (*fuzzy*) vrednosti kako bi se uz primenu jedne od mogućih funkcija dobila konkretna vrednost

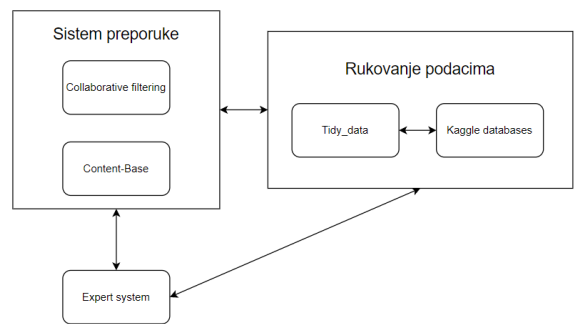


Slika 1 – Arhitektura fuzzy logike [5]

F. Arhitektura sistema

Arhitektura ovog sistema može se podeliti u tri celine (Slika 2):

- Rukovanje podacima
 - *Kaggle database* za pristup svim informacijama
 - *Tidy_data* za čišćenje i analizu podataka
- Sistemi za preporuku
 - *Collaborative filtering* za preporuku željene količine animea
 - *Content-Base filtering* proširenje prvobitne preporuke
- *Expert system* za finalnu preporuku



Slika 2 – Izgled arhitekture celokupnog sistema za preporuku

IV. REZULTATI I DISKUSIJA

A. Kaggle database:

Svi podaci koji se koriste u projektu su preuzeti sa kaggle sajta i to su:

- <https://www.kaggle.com/datasets/CooperUnion/anime-recommendations-database>
 - anime.csv sa osnovnim podacima o animeu
 - rating.csv sa informacijama o ocenama svakog korisnika za određeni anime.
- <https://www.kaggle.com/datasets/vishalmane10/anime-dataset-2022>
 - Anime.csv sa dodatnim podacima o animeima (konkretno tagovi su od najveće važnosti)

B. Tidy_data podsistem

Usled velike količine podataka neophonda je provera sadržaja koji se nalazi u tabelama. Za takav vid manipulacije podacima je korišćen *Tidy_data* podsistem. Pregledanjem kolone “*Genres*” detektovani su animei sa neprikladnim sadržajem i nepostojećim opisima. Obzirom da je njihova količina zanemarljiva u odnosu na ostatak podataka, oni su uklonjeni iz razmatranja u daljem razvoju projekta.

Kako bi se izbegli kasniji problemi u vidu preporuke različitih sezona istog animea, *Tidy_data* podsistem ima zadatak da odvoji grupe zajedničkih animea. Da bi se taj proces izvršio, neophodno je poređenje određenih podataka iz tabele anime.csv.

Poređenja su vršena koristeći funkciju sličnosti između dva animea (Slika 3). Atributi u odnosu na koje se animei porede jesu naziv, žanrovima i kreatori (kombinacija producenata i studijske kuće). Takva poređenja moraju sadržati različite koeficijente značajnosti. Razlog za uvođenje koeficijenata značajnosti jeste realna mogućnost da oba animea koja se porede poseduju identične žanrove i veoma slične informacije o kreatorima, ali da predstavljaju potpuno dva različita animea. Koeficijent koji predstavlja sličnost u nazivu animea mora da ima najveću vrednost. Konačne vrednosti koeficijenata (c1, c2, i c3) su dobijene eksperimentalnim putem (Slika 3). Za poređenje po nazivu smo koristili *Cosine Similarity* algoritam usled značaja ponavljanja reči u nazivu, a za žanrove i kreatore smo koristili *Jaccard Similarity* algoritam pošto je on izuzetno

brz, a ne postoji mogućnost ponavljanja vrednosti unutar tih kolona.

```
def similarity_score(anime1, anime2):
    c1, c2, c3 = 8, 1, 1
    name_cos_sim = cosine_similarity(anime1[Name], anime2[Name])
    creators_jac_sim = jaccard_sim(anime1[Creators], anime2[Creators])
    gen_jac_sim = jaccard_sim(set(anime1[Genres]), set(anime2[Genres]))
    score = c1 * name_cos_sim + c2 * creators_jac_sim + c3 * gen_jac_sim
    score = score / (c1 + c2 + c3)
    return round(score,3)
```

Slika 3 – Funkcija Similarity Score

Kao rezultat korišćenja matrice sličnosti dobijamo grupe zajedničkih animea (Slika 4) koje će se kasnije koristiti prilikom preporuke animea.

	Name	MAL_ID	Group
0	Bouken Ou Beet	8	0
1	Bouken Ou Beet Excellion	1123	0
2	Hachimitsu to Clover II	1142	1
3	Hachimitsu to Clover	16	1
4	Hachimitsu to Clover Specials	644	1
5	One Piece: Kinkyuu Kikaku One Piece Kanzen Kou...	16143	2
6	One Piece Recap	8171	2
7	One Piece	21	2
8	Tennis no Ouji-sama: Pair Prince	10862	3
9	Tennis no Ouji-sama	22	3
10	Tennis no Ouji-sama: Best GamesII	36952	3
11	Shin Tennis no Ouji-sama	11371	3
12	Shin Tennis no Ouji-sama Specials	13245	3
13	Tennis no Ouji-sama: Oshougatsu Special	5996	3

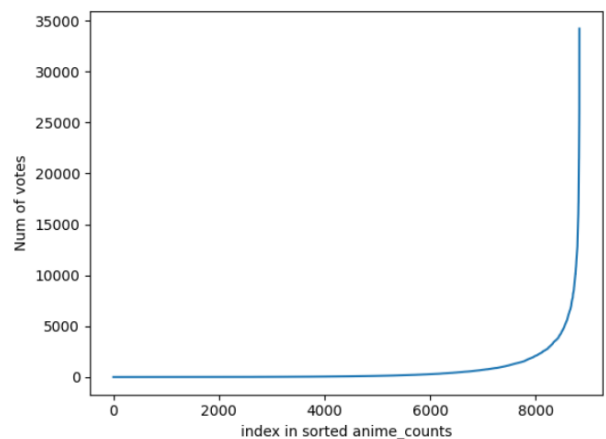
Slika 4 – Isečak iz novokreirane tabele grupa zajedničkih animea

Obzirom da ne postoji način da se sa sigurnošću testira uspešnost algoritma, rezultati su pregledani ručno i kreirane su zadovoljavajuće grupe.

C. Collaborative filtering

Collaborative filtering predstavlja algoritam preporuke koji iziskuje kreiranje ogromne, ali ujedno i retko popunjene *user-item* matrice.

Kako su resursi kojima rukujemo ograničeni, neophodno je smanjiti dimenzionalnost matrice. Podesni kandidati za izbacivanje su animei koji imaju mali broj glasova.



Slika 5 – Prikaz rastuće količine glasova za sve anime

Na vizuelnom prikazu glasova uočava se trend da najveći broj animea spada u grupu sa manjim brojem ocena (Slika 5). Daljom analizom odlučeno je da prag za najmanji broj glasova koji je neophodan da bi anime bio uvršten u *user-item* matricu bude deset glasova (Slika 6). Pod pretpostavkom da su takvi animei još uvek nedovoljno popularni, broj mogućih animea se smanjio sa 8826 na 6308 (28.49%). Na taj način je drastično smanjena dimenzionalnost *user-item* matrice koju ćemo kasnije kreirati.

br glasova	broj animea
1	800
2	482
3	293
4	226
5	145
6	151
7	120
8	106
9	90
10	105
11	64
12	64
13	70
14	79
15	77
16	61
17	58
18	43
19	47
20	41

Slika 6 – Tabela prikaza broja anime filmova sa ukupnim brojem glasova

Za kreiranje SVD (*Singular Value Decomposition*) modela preporuke od preostalih animea za razmatranje izvršena je podela na trening i test skup u odnosu 75% : 25%. Do takvog odnosa smo došli primenom *cross_validation* funkcije (Slika 7). SVD model kreira *user-item* matricu i nju dekomponuje na tri matrice – matricu korisnika, matricu proizvoda i matricu faktora. Koristeći 3 novokreirane matrice vrši se predikcija ocene koju bi korisnik dodelio za navedeni anime.

Evaluating RMSE, MAE of algorithm SVD on 5 split(s).

	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean	Std
RMSE (testset)	1.2366	1.2332	1.2337	1.2378	1.2348	1.2352	0.0017
MAE (testset)	0.9345	0.9322	0.9313	0.9332	0.9335	0.9329	0.0011
Fit time	35.22	28.23	30.78	27.20	32.13	30.71	2.86
Test time	5.72	5.02	5.00	5.77	5.95	5.49	0.40

Slika 7 – cross_validation funkcija

RMSE greška iznosi približno 1.2352 (*Slika 7*) i ona predstavlja grešku prilikom predviđanja ocene koju bi korisnik dodelio animeu. Takva greška je izuzetno zadovoljavajuća pošto su dodeljene ocene u rasponu od 1 do 10.

Nakon obučavanja modela, za konkretnog korisnika se preporuka izvršava tako što se koristeći SVD model, dodeljuje ocena svakom neocenjenom animeu. Animei sa predviđenim ocenama se sortiraju u opadajućem redosledu na osnovu dodeljene ocene i rezultat od prvih 25 preporuka se prosleđuje Content-Base sloju (*Slika 8*).

	MAL_ID	Name	Estimation
0	820	Ginga Eiyuu Densetsu	9.594884
1	199	Sen to Chihiro no Kamikakushi	9.517046
2	918	Gintama	9.444418
3	6594	Katanagatari	9.254977
4	431	Howl no Ugoku Shiro	9.215057
5	11979	Mahou Shoujo Madoka★Magica Movie 2: Eien no Mo...	9.085566
6	263	Hajime no Ippo	9.032332
7	164	Mononoke Hime	9.021596
8	1453	Maison Ikkoku	8.996126
9	232	Cardcaptor Sakura	8.981206
10	486	Kino no Tabi: The Beautiful World	8.914218
11	2418	Stranger: Mukou Hadan	8.887235
12	237	Koukyoushihen Eureka Seven	8.887190
13	1604	Katekyo Hitman Reborn!	8.887064
14	392	Yuu☆Yuu☆Hakusho	8.861298

Slika 8 – Slika isečka liste preporuke animea

Potrebno je još naglasiti da prilikom generisanja liste preporuke, u slučaju da se pojavi anime iz iste grupe sličnih animea (opisano u *tidy_data*), on se zanemaruje.

D. Content-Base filtering

Content-Base predstavlja algoritam koji se zasniva na analiziranju sadržaja animea. Da bi se obezbedio korisniji opis sadržaja animea uključena je dodatna tabela sa podacima o animeu i to konkretno informacije o tagovima koji predstavljaju proširenje žanrova (*Slika 9*). Usled uočavanja duplikata tagova nastalih lošim formatiranjem csv tabele bilo je neophodno srediti te podatke pre spajanja sa drugom tabelom. Obzirom na to da duplikati imaju izuzetno veliku sličnost u nazivu, koristili smo *Levenshtein distance* algoritam za njihovo detektovanje. Neophodno je da korisnik pregleda tabelu duplikata i ručno oblikuje i dodeli vrednost koloni *GroupName*, kako bi se te modifikacije primenile na kompletnoj tabeli sa animeima (*Slika 10*).

'15th Century', '16th Century', '17th Century', '18th Century', '19th Century', 'Abstract', 'Acting', 'Action', 'Adult Couples', 'Adult Industry', 'Adventure', 'Africa', 'Afterlife', 'Age Gap', 'Age Transformation', 'Aging', 'Agriculture', 'Alcohol', 'Alice in Wonderland', 'Aliens', 'All-Boys School', 'All-Girls School', 'America', 'Amnesia', 'Amusement Park', 'Ancient China', 'Androids', 'Androphobia', 'Angels', 'Animal Abuse', 'Animal Characteristics', 'Animal Protagonists', 'Animal Transformation', 'Anime Bancho', 'Anime Industry', 'Anime Tamago', 'Anime no Chikara', 'Anime no Me', 'Animeism', 'Anthropomorphic', 'Apartment Life', 'Apocalypse', 'Apprenticeship', 'Archery', 'Arranged Marriage', 'Art', 'Art School', 'Assassins', 'Astronauts', 'Astronomy', 'Australia', 'Autobiographies', 'Aviation', 'BDSM', 'BL', 'Badminton', 'Baking', 'Ballet Dancing', 'Bar', 'Baseball', 'Based on a 4-Koma Manga', 'Based on a Card Game', 'Based on a Cartoon', 'Based on a Comic Book', 'Based on a Doujinshi', 'Based on a Fairy Tale', 'Based on a Light Novel', 'Based on a Manga', 'Based on a Mobile Game', 'Based on a Movie', 'Based on a Novel', 'Based on a Picture Book', 'Based on a Play', 'Based on a Religious Text', 'Based on a Song', 'Based on a TV Series', 'Based on a Video Game',

Slika 9 – Isečak jedinstvenih vrednosti tagova

	Names:	GroupName:
0	{'Violence', 'Violence'}	Violence
4	{'Explicit Violence', 'Explicit Violence,'}	Explicit Violence
10	{'Mature Themes', 'Mature Themes'}	Mature Themes
12	{'Suicide', 'Suicide'}	Suicide
16	{'Cannibalism', 'Cannibalism,'}	Cannibalism
19	{'Bullying', 'Bullying'}	Bullying

Slika 10 – Isečak tabele duplokata

Problem koji nastaje spajanjem nove tabele sa već postojećom tabelom animea jeste što ne postoji jedinstven identifikator koji može biti korišćen. Neophodno je spajanje vrednosti tabele na osnovu naziva animea. Tabele koje se spajaju potiču iz različitih izvora i zbog toga je bitno naglasiti da se nazivi prvo normalizuju, kako bi se smanjila mogućnost greške. Za poređenje normalizovanih naziva korišćen je *Levenshtein distance* algoritma. Poređenjem se traže parovi sa najvećim procentom poklapanja i ukoliko on prelazi minimalni prag od 90%, smatramo da su dva animea identična.

Nakon što imamo sve neophodne informacije o saržaju animea izvršavamo *content-base* algoritam nad listom preporučenih animea, dobijenih iz prethodnog *Collaborative* sloja. Nad njima pronalazimo koeficijent najveće sličnosti sa sadržajem već ocenjenih animea od strane istog korisnika iz tabele *rating.csv*. Prikaz rezultata *Content-Base* sloja se oslikava u koloni *Similarity* (*Slika 11*). Sličnost se računa nad kolonom tagovi uz primenu *Term Frequency — Inverse Document Frequency* (TF-IDF) tehnike [6].

	MAL_ID	Name	Estimation	Similarity
0	820	Ginga Eiyuu Densetsu	9.594884	0.694365
1	199	Sen to Chihiro no Kamikakushi	9.517046	1.000000
2	918	Gintama	9.444418	0.614572
3	6594	Katanagatari	9.254977	1.000000
4	431	Howl no Ugoku Shiro	9.215057	1.000000
5	11979	Mahou Shoujo Madoka★Magica Movie 2: Eien no Mo...	9.085566	1.000000
6	263	Hajime no Ippo	9.032332	0.841716
7	164	Mononoke Hime	9.021596	1.000000
8	1453	Maison Ikkoku	8.996126	0.505286
9	232	Cardcaptor Sakura	8.981206	1.000000
10	486	Kino no Tabi: The Beautiful World	8.914218	1.000000
11	2418	Stranger: Mukou Hadan	8.887235	1.000000
12	237	Koukoushihen Eureka Seven	8.887190	0.843979
13	1604	Katekyo Hitman Reborn!	8.887064	0.560464
14	392	Yuu☆Yuu☆Hakusho	8.861298	1.000000

Slika 11 – Preporuka dobijena iz Collaborative sloja uz dodatak kolone sličnosti (Similarity)

E. Expert System

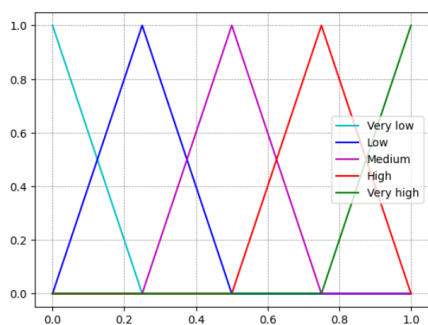
Expert system je poslednji element u arhitekturi sistema koji treba da dodeli finalnu odluku o redosledu preporuka. Zasniva se na fuzzy logici koja kalkuliše značaj animea (*EXS IMPORTANCE*) na osnovu tri ulazne vrednosti:

- sličnosti animea iz content-base sloja (*Similarity*)
- prosečna ocena (*Score*)
- ukupan broj glasova (*total_votes*)

1) Ulazne varijable

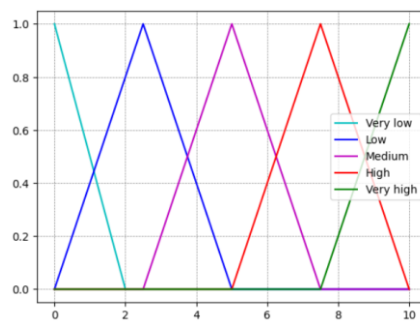
Sva tri parametra predstavljaju ulazne varijable fuzzy sistema opisane vrednostima *very_low*, *low*, *medium*, *high*, *very_high*. Funkcije pripadnosti za ovakve ulazne varijable su sledeće:

a) *similarity* – vrednosti opsega od 0 do 1 (Slika 12)



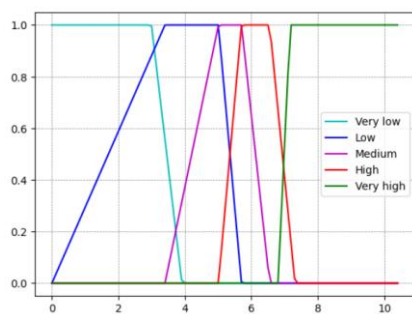
Slika 12 – trouglasta funkcija pripadnosti similarity

b) *score* – vrednosti opsega od 1 do 10 (Slika 13)



Slika 13- trouglasta funkcija pripadnosti score

c) *total_votes* – vrednosti opsega od 10 do 35000 (Slika 14)

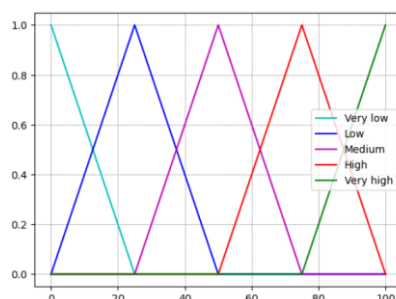


Slika 14 – trapezoidna funkcija pripadnosti total_votes

Jedino se za *total_votes* koristi trapezoidna funkcija opisa i do te odluke se došlo pošto je jako teško navesti u kojoj tački bi vrednost *total_votes* bila ekstrem za neku od navedenih opcija (*very_low*, *low*, ...). Za lepši prikaz pragova, korišćena je logaritamska funkcija. Do takve odluke se došlo zbog činjenice da se za *very_high* posmatra prag glasova od 1500 i tada bi on zauzimao oko 95% ukupnog grafika, čime ostali pragovi ne bi došli do izražaja.

2) Izlazna varijabla - importance

Za izlaznu varijablu, koja predstavlja rezultat *fuzzy* logike koristili smo takođe trouglastu funkciju koja je jednako raspoređena. Rešenje koje fuzzy logika vraća se nalazi u opsegu od 0 do 100%. (Slika 15)



Slika 15 - Troughlasta funkcija pripadnosti importance

3) Pravila - rules

Kako bi korišćenje *fuzzy* logike bilo moguće, neophodno je uvođenje pravila (*rules*) kao što je prikazano na *Slika 16*:

avg rating	num of votes	similarity	importance
low	low	low	very_low
low	low	very_high	low
medium	medium	very_high	medium
medium	high	very_high	medium
high	high	very_high	high
high	very_high	very_high	very_high
medium	very_high	very_high	high

Slika 16 – tabela pravila

Za uspešno izvršavanje logike bilo je neophodno navesti dvadesetak pravila. Rezultat koji dobijamo se upisuje u kolonu *EXS IMPORTANCE* (*Slika 18*) i nju koristimo u finalnoj evaluaciji (*final_evaluation*) kako bismo odredili konačan rezultat preporuke. *Final_evaluation* se računa prema formuli dobijenoj iz primera naučnog rada [3] (*Slika 17*):

```
def calculate_final_evaluation(x):
    est = x['Estimation'] / 10
    impo = x['EXS IMPORTANCE'] / 100
    if x['Estimation'] <= 0:
        return est * impo
    return est * (1+impo)
```

Slika 17 – pseudo kod za izračunavanje final_evaluation podatka

Nakon dobijene finalne evaluacije, rezultat već preporučenih animea sortiramo u opadajućem redosledu posmatrajući samo tu kolonu. Korisniku plasiramo prvih 15 preporuka (*Slika 18*).

	MAL_ID	Name	Estimation	Similarity	total_votes	Score	EXS_IMPORTANCE	Final_evaluation
1	199	Sen to Chihiro no Kamikakushi	9.517046	1.000000	22974	8.83	91.666667	1.824100
0	820	Ginga Eiyuu Denetsu	9.594084	0.694365	903	9.07	89.861817	1.821702
3	6594	Katanagatari	9.254977	1.000000	5317	8.36	91.666667	1.773871
4	431	Howl no Ugoku Shiro	9.215057	1.000000	17258	8.67	91.666667	1.766219
5	11979	Mahou Shoujo Madoka★Magica Movie 2: Eien no Mo...	9.085566	1.000000	3014	8.39	91.666667	1.741400
7	164	Mononoke Hime	9.021596	1.000000	16303	8.72	91.666667	1.729139
9	232	Cardcaptor Sakura	8.981206	1.000000	7452	8.16	91.666667	1.721398
10	486	Kino no Tabi: The Beautiful World	8.914218	1.000000	2550	8.34	91.666667	1.708558
2	918	Gintama	9.444418	0.614572	4974	8.96	80.783223	1.707392
11	2418	Stranger: Mukou Hadan	8.887235	1.000000	5115	8.31	91.666667	1.703387
8	1453	Maison Ikkoku	8.996126	0.505286	438	8.18	89.110802	1.701265
14	392	Yuu☆Yuu☆Hakusho	8.861298	1.000000	6764	8.45	91.666667	1.698415
16	721	Princess Tutu	8.823084	1.000000	2100	8.15	91.666667	1.691091
17	572	Kaze no Tani no Nausicaa	8.819509	1.000000	7109	8.4	91.666667	1.690406
19	122	Full Moon wo Sagashite	8.804714	1.000000	2103	7.96	91.666667	1.687570

Slika 18 – prikaz konačnog rezultata preporuke expert sistema

Na osnovu prve kolone indeksa, može se uočiti kako je *Expert* sistem izvršio uticaj na redosled preporuka iz početne liste koju je generisao *Collaborative* sloj (*Slika 18*).

V. ZAKLJUČAK

Ovim radom je postignuto da korisnik dobije preporuku ne samo na osnovu grupe sličnih korisnika, nego i njegovih ličnih preferencija u vidu tagova animea.

Mogućnosti daljeg rada na ovom projektu su poboljšanje algoritma za odvajanje grupa istih animea. Trenutno on funkcioniše po principu da ako bar jedan član grupe ima sličnost koja prelazi definisani prag sa novim animeom, novi anime će biti ubačen u istu grupu. Način da se prevaziđe taj problem jeste dodatak obavezne kontrole sličnosti novog animea sa celokupnom grupom. Mogućnosti za napredak se takođe uočavaju unutar *Content-Based* sloja, tako što ćemo proširiti parametre koji ulaze u kalkulaciju, ne samo tagove.

Konačno uz dodatno istraživanje *fuzzy* logike moguće je proširiti pravila, kao i modifikovanje ulaznih i izlaznih varijabli i funkcija kako bi se došlo do boljih rezultata. *Expert* sloj nam ostavlja mogućnost i za korišćenje neuronskih mreža kao vid načina preporuke.

REFERENCES

- [1] Isinkaye, F., Folajimi, Y., & Ojokoh, B. (2015). Recommendation systems: principles, methods and evaluation. *Egyptian Informatics J* 16 (3): 261–273.
- [2] Reddy, S. R. S., Nalluri, S., Kuniseti, S., Ashok, S., & Venkatesh, B. (2019). Content-based movie recommendation system using genre correlation. In *Smart Intelligent Computing and Applications: Proceedings of the Second International Conference on SCI 2018, Volume 2* (pp. 391-397). Springer Singapore.
- [3] Walek, B., & Fojtik, V. (2020). A hybrid recommender system for recommending relevant movies using an expert system. *Expert Systems with Applications*, 158, 113452.
- [4] Python, R. (n.d.). Build a Recommendation Engine With Collaborative Filtering – Real Python. Build a Recommendation Engine With Collaborative Filtering – Real Python. <https://realpython.com/build-recommendation-engine-collaborative-filtering/>
- [5] Fuzzy Logic | Introduction - GeeksforGeeks. (2018, April 10). GeeksforGeeks. <https://www.geeksforgeeks.org/fuzzy-logic-introduction/>
- [6] Scott, W. (2021, September 26). TF-IDF for Document Ranking from scratch in python on real world dataset. Medium. <https://towardsdatascience.com/tf-idf-for-document-ranking-from-scratch-in-python-on-real-world-dataset-796d339a4089>