Zbornik radova Fakulteta tehničkih nauka, Novi Sad

UDK: 004.85
DOI: https://doi.org/10.24867/21BE10Jovanovic

PREDIKCIJA ŽANRA PESME NA OSNOVU TEKSTA PESME POMOĆU ALGORITAMA MAŠINSKOG UČENJA

PREDICTION OF SONG GENRE BASED ON SONG LYRICS USING MACHINE LEARNING ALGORITHMS

Marko Jovanović, *Fakultet tehničkih nauka, Novi Sad*

Oblast – RAČUNARSTVO I AUTOMATIKA

Kratak sadržaj – *Muzički svet je u procesu eksponencijalnog rasta, naročito u proteklih desetak godina kada se zahvaljujući digitalizaciji dogodila značajna ekspanzija. Broj novih umetnika i novih pesama je u konstantnom porastu, što je dovelo do potrebe za kreiranjem inteligentnog i efikasnog načina da se slušaoci snađu u moru različitih izbora. Jedan od kvalitetnih načina za filtriranje numera je filtriranje na osnovu žanra pesme. U ovom radu će biti opisan sistem za klasifikaciju pesama u različite žanrove na osnovu teksta pesme. Podaci za potrebe obučavanja i testiranja sistema su prikupljeni iz više različitih izvora i transformisani tako da čine jedan jedinstven skup podataka. Najbolji rezultati su postignuti upotrebom GloVe vektorskih reprezentacija reči i* LSTM *arhitekture mreže. Postignuta je tačnost od 78%, preciznost od 78%, odziv od 77% i* F-*mera od 77%.*

Ključne reči: *mašinsko učenje u muzici, duboko učenje u muzici, predikcija žanra pesme*

Abstract – *The world of music is in the process of exponential growth, especially in the past ten years when a significant expansion took place thanks to the digitalization of the world. The number of new artists and songs is constantly increasing, which has led to the need to create an intelligent and efficient way for listeners to navigate the myriad of choices. One of the proven ways to filter tracks is to filter based on song genre. This paper will describe a system for classifying songs into different genres based on the song's lyrics. Data for training and testing the system is collected from several different sources and transformed into a single data set. The best results were achieved using GloVe word embeddings and LSTM network architecture. This model achieved 78% accuracy, 78% precision, 77% recall, and 77% F-measure.*

Keywords: *machine learning in music, deep learning in music, song genre prediction*

1. UVOD

Automatska klasifikacija muzike po žanrovima jedan je od glavnih zadataka u oblasti *music information retrieval*-a (MIR). Popularizacijom mobilnih telefona i drugih prenosivih uređaja, platforme za reprodukciju muzike kao što su *Spotify* [1] i *Apple music* [2] dobijaju na značaju i konstantno se takmiče u pokušajima da svojim korisnicima pruže kvalitetne muzičke preporuke.

Kvalitet preporučene pesme mnogo zavisi od raspoloživih meta-podataka o pesmi. Generisanje meta-podataka o pesmama je u prošlosti bio manuelni zadatak i često je veoma dugo trajao. Kako je sve više sadržaja dostupno na internetu, a i dinamika konzumiranja sadržaja postaje sve kompleksnija, moderna rešenja su postala potrebna kako bi anotacija numera bila kvalitetna i brza. Razvojem različitih algoritama mašinskog učenja, kao i razvojem samog hardvera računara, omogućeno je obučiti različite modele koji će automatski obavljati posao generisanja meta-podataka o pesmama.

U ovom radu su, za rešavanje problema automatske klasifikacije pesme po žanrovima na osnovu teksta pesme, korišćene tehnike obrade prirodnog jezika (engl. *Natural Language Processing* - NLP). Prilikom rešavanja NLP problema, cilj je da se tekstu pesme pridruže odgovarajuće labele sa semantičkim značenjem, koje u ovom slučaju predstavljaju žanr pesme.

Moderna rešenja za klasifikaciju tekstualnog sadržaja su najbolje opisana u radu [6]. Sva moderna rešenja, na kojima se zasnivaju modeli opisan u ovom radu, se temelje na kompleksnim vektorskim prostorima za reprezentaciju reči, dubokom učenju i na arhitekturama konvolucionih neuronskih mreža i LSTM mreža.

Glavni cilj ovog rada je da detaljno opiše proces pripreme podataka da bi bili u odgovarajućem formatu za uspešno treniranje modela dubokog učenja. Takođe, predstavljeni su različiti algoritmi dubokog učenja kojima je vršena predikcija žanra pesme na osnovu teksta pesme. Na ovaj način se obogaćuje skup meta-podataka i rešava se prethodno opisan problem. Sistem predstavljen u radu uspešno uočava različite skupove reči karakteristične za pojedine žanrove i pomoću tih informacija klasifikuje pesme u jedan od deset mogućih žanrova. Pokazano je da zbog prirode problema i teškoće učenja pravila u govornom jeziku kompleksniji modeli dubokog učenja postižu znatno bolje rezultate.

2. PRETHODNA REŠENJA

Slični radovi su odabrani tako što je uzeta u obzir metodologija i slični modeli mašinskog učenja. Prethodna rešenja su takođe rešavala problem automatske klasifikacije žanra pesme na osnovu teksta pesme.

NAPOMENA:
Ovaj rad proistekao je iz master rada čiji mentor je bila dr Jelena Slivka, vanr. prof.

Takođe, uzeti su u obzir radovi koji koriste skupove podataka generisane na različite načine.

Upotreba hijerarhijskih mreža sa mehanizmom pažnje (engl. *Hierarchical Attention Network* - HAN) je predložena u radu [3]. HAN uzimaju u obzir unutrašnju strukturu i organizaciju podataka kako bi se unapredio model. Primećeno je da se muzičke numere sastoje od reči koje čine segmente, a segmenti čine pesmu u celosti. HAN se primenjuje za klasifikaciju pesme na osnovu više pomenutih nivoa teksta pesme, počev od nivoa reči, do nivoa segmenta. Cilj rada [3] je bio da se pokaže da ovakva arhitektura pruža bolje rezultate u odnosu na arhitekture koje ne uzimaju u obzir hijerarhijsku organizaciju. Skup podataka koji je korišćen u radu [3] predstavlja kolekciju tekstova pesama prikupljenih upotrebom licenciranog softvera *LyricFind* [4]. Čitav skup podataka se sastojao od preko milion pesama koje su bile sačuvane u JSON formatu. Pošto u ovom skupu podataka nisu bile dostupne informacije o žanru pesme, početni skup podataka je proširen. Informacije o žanru pesme su preuzete pomoću aplikacije *iTunes Search API* [5]. Iz skupa podataka su izbačeni oni žanrovi koji su imali manje od 50 pesama predstavnika tog žanra. Ove transformacije su dovele do toga da se finalni skup podataka sastojao od oko pola miliona tekstova pesama koje su pripadale jednom od 20 različitih žanrova. Obučena neuronska mreža sa mehanizmom pažnje je postigla preciznost od 43.66%.

Jedan od zanimljivijih radova na ovu temu bio je rad [6] koji poredi više različitih algoritama mašinskog učenja i njihove rezultate. Korišćeni su model nasumične šume (engl. *random forest*), kao i konvolucione i LSTM arhitekture dubokih neuronskih mreža. Opisani su samo najbolji rezultati koji su postignuti LSTM arhitekturom mreže. Broj žanrova za predikciju je bio sedam. Tačnost za ovaj pristup je iznosila 62.4%.

Rad [7] se zasniva na pretpostavci da se tekstovi pesama dovoljno razlikuju da bi automatska detekcija žanra pesme, kvaliteta pesme, kao i godine izdavanja pesme bili mogući. Pokazano je da klasifikatori koji uzimaju u obzir informacije o tekstu pesme pružaju kvalitetnije predikcije od onih koji koriste samo zvuk pesme. Pored detekcije žanra, u radu [7] je predstavljeno i rešenje koje razlikuje dobre i loše pesme. Zbog nedostatka dovoljno velikog skupa podataka autori su koristili tekstove sa *LyricsMode* [8] veb stranice zbog velike pokrivenosti različitih žanrova i konzistentnosti u podacima. Primenjene su metode pretprocesiranja teksta, uklanjanje duplikata i tokenizacija reči pesama. U finalni skup podataka su uključene samo pesme koje su na engleskom jeziku. Konačni skup podataka se sastojao od 400 hiljada tekstova pesama i pratećih žanrova. Za klasifikaciju je korišćen *n-gram* model, a broj žanrova za predikciju je bio osam.

3. SKUPOVI PODATAKA

Za potrebe ovog rada, podaci su prikupljeni iz više različitih izvora. Informacije o imenu pesme i žanrovima su preuzete iz skupa podataka *Million Song Dataset* - MSD [9], dok su informacije o tekstovima pesme preuzete sa *Genius* [10] API-ja. Proces preuzimanja tekstova pesme je bio iterativan. Za svako ime pesme i ime izvođača iz skupa MSD poslat je zahtev na *Genius* API koji je vraćao tekst tražene pesme. Ukoliko nekim slučajem tekst pesme nije bilo moguće pronaći, pesma bi bila izbačena iz skupa podataka. Dešavali su se slučajevi gde je za određenu pesmu postojalo više mogućih tekstova pesme. Ovaj problem je rešen tako što je uzet u obzir samo prvi predloženi tekst pesme, a svi ostali su odbačeni.

Za pretprocesiranje podataka korišćene su različite tehnike za obradu prirodnog jezika. Prvo su svi karakteri prebačeni u mala slova. Nakon toga su izbačeni znaci interpunkcije. Da bi se uklonile nevažne reči u tekstovima iskorišćen je predefinisani skup stop reči iz NLTK [11] biblioteke. Lematizacijom su reči prebačene u svoj osnovni oblik. Na kraju su sve reči predstavljene pomoću vektorskih reprezentacija. Algoritmi koji su korišćeni za ovaj zadatak su bili *Word2Vec* i *GloVe* [12] .

Da bi se model mogao obučiti, da bi se podesili hiper-parametri i da bi se evaluirao model, neophodno je imati tri skupa podataka. Trening skup podataka od 80% početnog skupa podataka je korišćen za obučavanje modela, validacioni skup od 10% početnog skupa je korišćen za podešavanje hiper-parametara, a test skup koji sadrži 10% početnog skupa podataka je korišćen za evaluaciju rešenja. Zbog kompleksnosti korišćenih modela i dugačkog trajanja obučavanja različitih modela nije iskorišćena višestruka unakrsna validacija koja bi verovatno bila bolji način za pronalaženje optimalnih vrednosti hiper-parametara.

4. METODOLOGIJA

U ovom poglavlju će biti opisana arhitektura modela za predikciju žanra pesme kroz tri poglavlja: (1) vektorske reprezentacije reči, (2) konvolucione neuronske mreže, (3) LSTM arhitektura konvolucione neuronske mreže. Takođe, u odvojenom poglavlju će biti opisan proces obučavanja modela.

4.1. Vektorske reprezentacije reči

Word2Vec je algoritam pomoću kog se kreiraju vektorske reprezentacije reči. Algoritam je zasnovan na modelu neuronske mreže za učenje asocijacija reči iz velikog korpusa teksta. Ovako kreirane reprezentacije reči uzimaju u obzir i kontekst, tako da su sliče reči ili sinonimi bliže u vektorskom prostoru.

Pretrenirane *GloVe* vektorske reprezentacije reči su dotrenirane nad postojećim korpusom. Broj dimenzija je bio podešen na 100, dok je veličina rečnika iznosila 400 hiljada reči.

4.2. Konvoluciona neuronska mreža

Za konstrukciju arhitekture konvolucione neuronske mreže korišćena je *Keras* [13] biblioteka.

Funkcija gubitka koja se koristila je unakrsna entropija (engl. *cross entropy*). Unakrsna entropija se tipično koristi kod problema klasifikacije. *Adam* je korišćen kao optimizacioni algoritam stohastičkog gradijentnog spusta. Aktivaciona funkcija na izlazu je bila *SoftMax* i klasifikacija se vršila za desert mogućih žanrova.

Ulaz u mrežu je bio 2D niz ograničen na dužinu od 2000 vektorskih reprezentacija reči u tekstu pesme, uključujući i *padding*. Svaka reč je predstavljena odgovarajućom vektorskom reprezentacijom čiji je broj dimenzija 100. *Pooling* sloj nije korišćen jer je bitna pozicija reči u tekstu za predikciju žanra.

Izlaz konvolucione neuronske mreže je vektor od desert vrednosti. Svaka vrednost predstavlja verovatnoću pripadanja određenom muzičkom žanru

4.3. LSTM arhitektura mreže

Za potrebe treniranja i konstrukcije arhitekture LSTM neuronskih mreža korišćena je *Keras* [13] biblioteka. Ova biblioteka je upotrebljena zbog velike fleksibilnosti i lakoće kreiranja različitih modela neuronskih mreža.

Kao funkcija gubitka korišćena je unakrsna entropija. Za potrebe algoritma stohastičkog gradijentnog spusta korišćen je *Adam* optimizacioni algoritam. Ovaj algoritam karakteriše adaptivna stopa učenja. Maksimalna dužina sekvence reči podešena je na 2000, dok je broj dimenzija vektora reprezentacije reči bio 100. Aktivaciona funkcija koja se koristila u poslednjem sloju je *SoftMax* funkcija, koja je pogodna za rešavanje problema klasifikacione prirode.

Cilj, odnosno izlaz, ove neuronske mreže je predikcija odgovarajućeg žanra pesme na osnovu ulaza koji predstavlja sekvencu vektorskih reprezentacija reči. Izlazna vrednost je vektor od desert vrednosti verovatnoće pripadanja svakom od žanrova.

4.4. Obučavanje modela

Za obučavanje, validaciju i testiranje modela za klasifikaciju žanra pesme korišćen je hibridni skup podataka. Ime pesme, ime izvođača, žanr i tekst pesme su preuzeti iz skupa podataka sa *Million songs dataset* [9] i *Genius* [10] API-ja. Skup podataka je sadržao oko million različitih numera i meta-podatke o njima. Inicijalno je postojalo 60 različitih žanrova, ali su modeli obučavani na podskupu od 10 najrelevantnijih žanrova.

Kod obučavanja modela konvolucione neuronske mreže broj epoha je bio podešen na 102, iz razloga što je za svako treniranje bilo potrebno ~14 časova da se završi. *Batch size* hiper-parametar je bio podešen ja 64.

Obučavanje LSTM arhitekture mreže je bio veoma slično kao i obučavanje konvolucione neuronske mreže. Broj epoha je bio podešen na 102, dok je *batch size* podešen na 64.

5. EVALUACIJA REŠENJA I REZULTATI

Mere evaluacije sistema za klasifikaciju pesama korišćene u ovom radu su F1-mera, tačnost, odziv i preciznost. Za računanje ovih metrika neophodno je imati poznate vrednosti iz matrice konfuzije (engl. *confusion matrix*) [14].

Tačnost se računa kao količnik ukupnog broja tačnih predikcija i ukupnog broja predikcija:

$$Tačnost = \frac{broj\ tačnih\ predikcija}{ukupan\ broj\ predikcija}.$$

Odziv daje informaciju koliko stvarno pozitivnih primera je model uspeo tačno da predvidi i računa se:

$$Odziv = \frac{stvarno\ pozitivni}{stvarno\ pozitivni + lažno\ negativni}.$$

Preciznost predstavlja meru koliko je model tačan pri davanju procene da je neki od primera pozitivan u odnosu na sve primere koji su procenjeni kao pozitivni:

$$Preciznost = \frac{stvarno\ pozitivni}{stvarno\ pozitivni + lažno\ pozitivni}.$$

F1-mera se računa kao kombinacija mere odziva i preciznosti:

$$F1mera = 2 * \frac{preciznost * odziv}{preciznost + odziv}.$$

Za svaki od obučenih modela izračunate su vrednosti za tačnost, odziv, preciznost i F1-meru. Tabela 1 prikazuje rezultate primene:

- Kombinacije konvolucionih neuronskih mreža i *GloVe* vektorskih reprezentacija reči
- Kombinacije LSTM arhitekture neuronske mreže i *Word2Vec* vektorskih reprezentacija
- Kombinacije LSTM arhitekture mreže i *GloVe* vektorskih reprezentacija reči

Tabela 1. *Performanse obučenih modela*

Naziv modela	Tačnost	Preciznost	Odziv	F-mera
CNN + GloVe	68%	64%	63%	62%
LSTM + Word2Vec	72%	72%	71%	70%
LSTM + GloVe	**78%**	**78%**	**77%**	**77%**

U tabeli 1 je istaknut u kom se nalaze rezultati modela koji je postigao najbolje vrednosti metrika za evaluaciju. Iako je korišćen manji skup podataka, postignuti su veoma dobri rezultati. Teško je bilo pronaći relevantan naučni rad sa kojim bi mogli da se direktno uporede dobijeni rezultati zbog činjenice da je početni skup podataka kreiran kombinacijom različitih tehnika za prikupljanje podataka.

U tabeli 2 se nalazi detaljni prikaz performansi LSTM modela i *GloVe* vektorskih reprezentacija reči po žanrovima. Neki od žanrova za koje je muzika često tačno klasifikovana pripadaju "hip hop", "metal" i "pop" žanrovima. Ova pojava se objašnjava činjenicom da postoji mnogo karakterističnih reči i načina izražavanja u pomenutim žanrovima. Ovo znatno olakšava klasifikaciju i samim tim su i rezultati bolji od očekivanih. Sa druge strane, muzika koja pripada "rege" žanru i "klasična muzika" su klasifikovane sa slabijom tačnošću. Razlog je

mali boj pesama predstavnika ovih žanrova i činjenica da su ove pesme često na stranom jeziku.

Tabela 2. *Detaljni prikaz performansi LSTM + GloVe modela*

Žanr	F1-mera
Bluz	67%
Klasika	62%
Kantri	79%
Disko	78%
Hip hop	93%
Džez	57%
Metal	91%
Pop	88%
Rege	56%
Rok	77%

6. ZAKLJUČAK

U ovom radu predstavljeno je i upoređeno više različitih pristupa za rešavanje problema automatske predikcije žanra pesme na osnovu teksta pesme. Pokazano je da, zbog prirode problema i teškoće učenja pravila u govornom jeziku, kompleksniji modeli dubokog učenja postižu bolje rezultate. Takođe, opisan je i proces pripreme podataka da bi bili u odgovarajućem formatu za uspešno treniranje modela dubokog učenja. Dat je i uvid u arhitekturu različitih modela koji su bili obučavani. Iz svega priloženog se može zaključiti da se najbolji rezultati postižu primenom *GloVe* vektorskih reprezentacija reči i LSTM arhitekture neuronske mreže.

Iako se celokupan sistem može opisati kao uspešan, neki od narednih koraka za dalji razvoj i unapređenje postojećeg rešenja su opisani u nastavku.

Prva ideja bi bila da se ceo skup podataka poveća tako da uključuje više pisama iz slabije zastupljenih žanrova. Ovaj proces može biti i sintetičke prirode gde bi se obučili modeli mašinskog učenja koji bi generisali tekstove pesama određenih žanrova.

Druga ideja jeste da se uzmu u obzir i meta-podaci o samoj pesmi i da se na taj način robusnijim skupom podataka poboljšaju rezultati predikcije. Neke od dodatnih osobina koje bi mogle da se koriste su dužina trajanja pesme, podaci o instrumentima koji se koriste, kao i karakteristike samog audio zapisa poput glasnoće i opsega frekvencija.

Treća ideja jeste da se upotrebom moćnijeg hardvera isproba više različitih arhitektura neuronskih mreža. Na ovaj način bi bilo moguće pronaći optimalne hiper-parametre mnogo brže i efikasnije.

Četvrta ideja jeste da se pokuša sa primenom transformer arhitektura poput BERT-a i GPT-3.

7. LITERATURA

[1] Spotify https://www.spotify.com/ [datum pristupa 18.09.2022.]

[2] Medium https://www.apple.com/apple-music/ [datum pristupa 18.09.2022.]

[3] Alexandros T. 2017. *Lyrics-based music genre classification using a hierarchical attention network*

[4] LyricFind https://www.lyricfind.com/ [datum pristupa 18.09.2022.]

[5] iTunes Search API https://developer.apple.com/library/archive/documentation/AudioVideo/Conceptual/iTuneSearchAPI/index.html [datum pristupa 18.09.2022.]

[6] Ciao Luiggy R., et. al. 2019. *Combining Diverse Models for Lyrics-based Music Genre Classification*

[7] Michael Fell, Caroline Sporleder, *Lyrics-based Analysis and Classification of Music*

[8] Wu, Chuhan & Wu, Fangzhao & An, Mingxiao & Huang, Jianqiang & Huang, Yongfeng & Xie, Xing. (2019). NPA: Neural News Recommendation with Personalized Attention.

[9] Lyrics mode https://www.lyricsmode.com/ [datum pristupa 18.09.2022.]

[10] Million Song Dataset http://millionsongdataset.com [datum pristupa 19.09.2022.]

[11] Genius https://genius.com [datum pristupa 19.09.2022.]

[12] NLTK https://www.nltk.org/ [datum pristupa 19.09.2022.]

[13] GloVe: Global Vectors for Word Representation https://nlp.stanford.edu/projects/glove/ [datum pristupa 19.09.2022.]

[14] Keras https://keras.io/ [datum pristupa 19.09.2022.]

[15] Multi-class Classification: Extracting Performance Metrics From The Confusion Matrix https://towardsdatascience.com/multi-class-classification-extracting-performance-metrics-from-the-confusion-matrix-b379b427a872 [datum pristupa 19.09.2022.]

Kratka biografija:

Marko Jovanović rođen je 07.03.1997. godine u Novom Sadu, Republika Srbija. Upisao se na Fakultet tehničkih nauka, odsek Računarstvo i automatika 2016. godine. Diplomirao je 2020. godine i iste godine upisuje se na master akademske studije, odsek Računarstvo i automatika, smer Elektronsko poslovanje.