

Klasifikacija glazbe po žanru

Projektni prijedlog

Šime Batović Mislav Vučković
Andrija Mandić Marko Jukić

22. travnja 2020.

Uvodni opis problema

U ovom istraživanju baviti ćemo se klasifikacijom glazbe po žanrovima. Analiza pjesama i njihovo prepoznavanje su često istraživani problemi. Danas je dostupno mnogo aplikacija koje na temelju svega nekoliko sekundi snimke mogu prepoznati o kojoj se pjesmi radi. Naravno, područje istraživanja je i dalje u razvoju te su moguća poboljšanja u točnosti predviđanja. Naš problem će biti malo jednostavniji, pokušat ćemo na temelju 30 sekundi zvučnog zapisa pjesme predvidjeti njen žanr. Odabrani skup podataka dostupan je na Kaggle-u¹. Korišteni FMA dataset sadrži 8000 isječaka pjesama koje su ravnomjerno raspoređene u 8 glavnih žanrova. Postoji podjela na podžanrove koji se dalje dijele na još manje podpodžanrove. Zbog malog broja primjera po najsitnijoj granulaciji žanrova odlučili smo se za predikciju samo glavnih žanrova.

Cilj i hipoteze istraživanja problema

Cilj istraživanja je usporediti klasične modele problema klasifikacije u strojnom učenju s metodom dubokog učenja koristeći konvolucijske neuronske mreže. Ideja je provjeriti već dokazane metode u radu s klasifikacijom zvuka te analizom i promjenom/uvođenjem novih značajki poboljšati krajnju preciznost predviđanja.

¹<https://www.kaggle.com/aaronyim/fma-small>

Pregled dosadašnjih istraživanja

Analiza i klasifikacija glazbe u današnje doba je duboko istražen problem. Popularne aplikacije poput Spotify i Google Play Music se već dugi niz godina bave ovim problemom. Navest ćemo dva različita istraživanja na koja smo naišli.

Grzegorz i Grzywczak su u svom radu [4] 2014. godine analizirali poznati GTZAN skup podataka koji sadrži 1000 pjesama iz 10 žanrova u .au formatu. Postigli su točnost od 78% izvlačenjem značajki iz spektograma koristeći duboke konvolucijske mreže (CNN) trenirane za klasifikaciju slika. Konačno za klasifikaciju žanra su koristili SVM. Očiti nedostatak jest što se ulazne podatke promatralo kao obične slike, a ne kao reprezentaciju uzorka glazbe. Također, mi ćemo u našem skupu podataka koristiti uzorke u mp3 formatu što može utjecati na postignutu točnost.

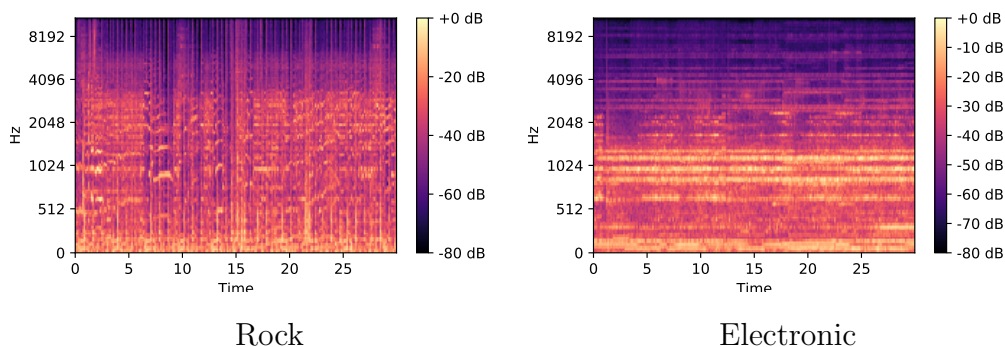
Još jedno istraživanje[1] na koje smo naišli koristilo je također GTZAN dataset od 1000 pjesama iz 10 žanrova. Uspoređivali su klasičan pristup izvođenja podataka koristeći SVM i Deep Learning pristup koristeći CNN (Convolution neural networks) na spektogramskom zapisu zvuka. U prvom pristupu koristili su razne modele bazirane na tempu i frekvenciji. Koristeći model slučajne šume postigli su točnost od 67.7%, a koristeći model stabla odlučivanja 51.6%. Koristili su još nekoliko modela te su to usporedili s modelom dubokog učenja koji su trenirali na skupu od 70% podataka i testirali na test skupu koristeći pristup glasa većine te dobili točnost od 83.2%.

Materijali, metodologija i plan istraživanja

Za skup podataka odabrali smo FMA dataset [3], konkretnije fma_small, koji se sastoji od 8000 isječaka pjesama duljine 30 sekundi u mp3 formatu. Uz svaku pjesmu dani su i brojni metapodaci, od kojih ćemo mi koristiti samo žanr. Iako je žanr dan kao hijerarhijski, mi ćemo zbog jednostavnosti i veličine skupa podakata pokušati predvidjeti samo jedan od osam glavnih žanrova: Experimental, Hip-Hop, Rock, Pop, Folk, Electronic, Instrumental i International.

Plan nam je ispitati i usporediti točnost klasičnih metoda strojnog učenja (stabla odlučivanja, slučajne šume, logistička regresija, SVM) i metoda dubokog učenja (CNN). Isprobat ćemo značajke iz prijašnjih istraživanja i provjeriti kako nove značajke utječu na točnost modela. Istražit ćemo i mogu li podaci o podžanrovima ipak pridonijeti točnosti predikcije.

Za implementaciju modela koristit ćemo Python i module scikit-learn², Keras³ i LibROSA⁴.



Slika 1: Primjer mel-spektograma dobivenih korištenjem modula LibROSA

Očekivani rezultati predloženog projekta

Dakle, kao što smo napomenuli: koristit ćemo konvolucijsku neuronsku mrežu, sa spektrogramima kao inputima, i jednu metodu strojnog učenja gdje ćemo sami izvlačiti atribute iz pjesama. Naš skup podataka je veći od GTZAN datasea korištenog u [1, 4], pa možemo očekivati bolje rezultate koristeći slične modele. Očekujemo da ćemo poboljšati točnost klasičnih metoda, modificiranjem/dodavanjem nekih atributa.

Literatura

- [1] Music genre classification on GTZAN dataset using CNNs. <https://github.com/Hguimaraes/gtzan.keras>.
- [2] Strojno učenje. <https://web.math.pmf.unizg.hr/nastava/su/>.
- [3] Michaël Defferrard, Kirell Benzi, Pierre Vandergheynst, and Xavier Bresson. FMA: A dataset for music analysis. In *18th International Society for Music Information Retrieval Conference*, 2017.
- [4] Daniel Grzywczak and Grzegorz Gwardys. Deep image features in music information retrieval. volume 60, pages 187–199, 08 2014.

²<https://scikit-learn.org/>

³<https://keras.io/>

⁴<https://librosa.github.io/>