

1. Uvodni deo

Klasifikacija audio podataka je jako bitna za funkcionisanje velikog broja aplikacija (Shazam , automatsko kucanje teksta na osnovu audia koje omogućava google ,automatsko pravljenje subtit-ova za filmove itd...)

Dafinicija problema: kako da klasifikujem audio signale u N vec poznatih kategorija

Moj projekat se oslanja na naučni rad

Music Genre Classification using Machine Learning Algorithms: A comparison

Cilj mog projekta je da za pesmu sa youtub-a prepozna kojem muzičkom žanru pripada

2. Opis vašeg rešenja zadatog problema

Za rešavanje ovog problema koristio sam potpuno povezane neuronske mreže (adam optimizer).

Arhitektura mreže: 4 sloja , 100 neurona u svakom sloju , relu I sigmoid aktivaciona funkcija.

Arhitektura za koju sam se odlučio je jako slična onoj iz naučnog rada

Koristio sam biblioteku keras I programski jezik Python.

Download-ovao sam 400 pesama sa jutjuba (100 heavy metal , 100 regge ,100 elektrosnkih pesama , 100 klasičnih kompozija).

Podelio sam taj skup pesama na TreningValidacioni I Testing skup u odnosu 9:1.

TreningValicacioni skup sam podelio na trening I validacioni skup u donosu 4:1.

Koristio sam unakrsnu validaciju radi preciznije ocene modela.

Svaki fajl pesme se može interpretirati kao jačina zvuka u određenom vremenskom trenutku

3. Poglavlje sa eksperimentalnim rezultatim

Pokušao sam da pristupim rešavanju ovog problema na 3 različita načina

1. Naivni pristup

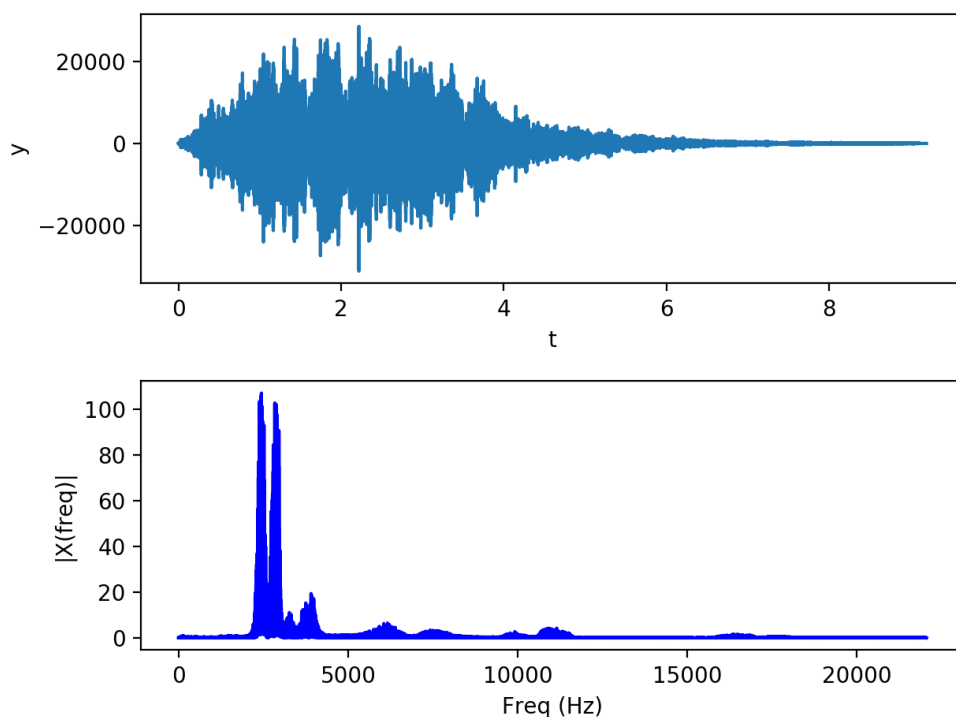
Prosledio sam podatke iz audiofajlova neizmenjene u neuronsku mrežu u nadi da će tačno klasifikovati makar dobar deo podataka , na žalost preciznost je bila 25% .

Naivan pristup se ne koristi u literaturi I sad znam zašto.

2. Način koji se primenjuje u naučnoj literaturi

Dobar deo naučnih radova preporučuje da se klasifikacija vrši na osnovu dominantnih frekvencija iz audio fajla.

Da bi se dobila učestalost svake frekvencije vrši se furijeva transformacija originalnog fajla.



Napisao sam program koji vrši furijevu transformaciju i vraća python listu od 25 000 vrednosti. svaka vrednost predstavlja učestalost te frekvencije u hercima.

Pre prosledjivanja audio podataka mreži (25 000 je prevelik input i sporo će se trenirati model) odlučio sam da diskretizujem (binujem) podatke

Za treniranje mreže odlučio sam se za sledeći algoritam

1. Trenira se prva epoha na trening podacima
2. Upoređuje se preciznost na trening i validacionom skupu
3. Ako je preciznost manja na validacionom $\text{BrojPutakojiSmeDaPogresi} = 1$
4. Ako je $\text{BrojPutakojiSmeDaPogresi} == 0$ prekida se treniranje

$\text{BrojPutakojiSmeDaPogresi} = 3$

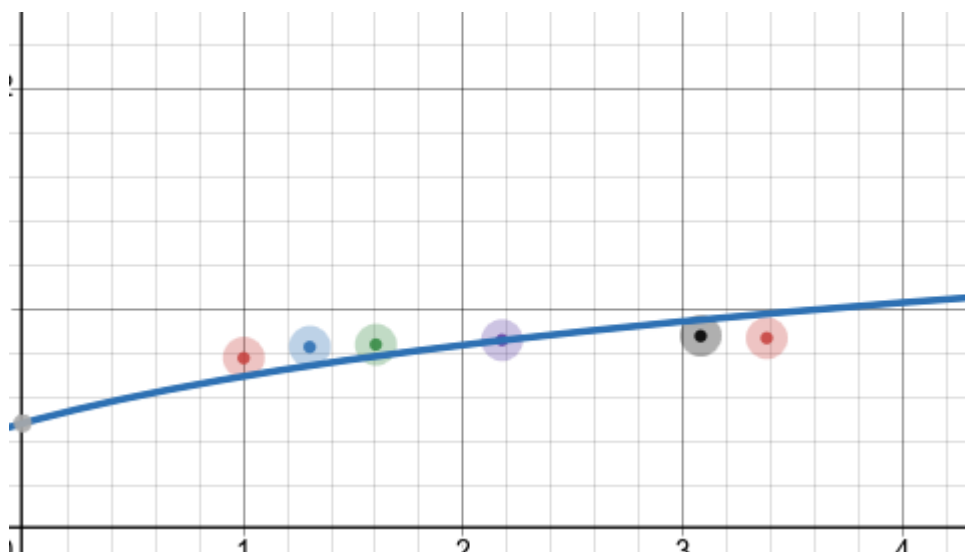
Mislim da je 3 korektna početna vrednost

Ekspreimatisao sam sa veličinom bina I upoređivao rezultate dobijene na validacionom skupu

Broj binova	Preciznost na validacionom
10	75%
20	79%
300	88%
600	87%
1200	83%

Preciznost je primetno rasla , ali samo do neke tačke

Vizualizovao sam ovo koristeći logističku regresiju



3. Unapređenje

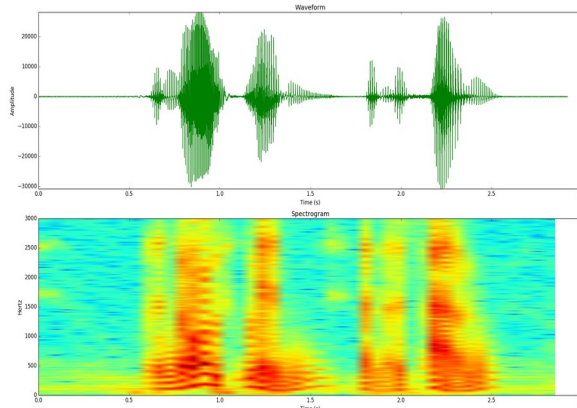
U naučnoj literaturi se spomnjao još jedan metod klasifikacije podataka.

Ideja je da se umesto jedne furejeve transformacije na celom fajlu on podeli na k manjih delova I da se za svaki mali delić uradi furijeve transformacija. Intuitivno ovo bi pomoglo da se klasifikuju žanrovi koji imaju karakteristične frekvencije na početku pesme.

Podelio sam svaku pesmu na 250 delova I uruadio furijeve transformaciju na njima , binovao sam po veličini 10 , dobio sam ulazne fajlove veličine 2500.

Tačnost na validacionom skupu je bila 46%.

Jedini zaključak koji mogu da izvučem je da žanrovi koje sam ja odabrao nemaju karakteristične početke I da sam ovime samo zbunio svoju neuronsku mrežu



Ovaj tip podatka se u literaturi naziva spektrogram

U naučnoj literaturi se spominje da su više frekvencije bitnije za klasifikaciju audio sadržaja (oni su odbacivali sve preko određene frekvencije)

Odlučio sam da testiram ovu hipotezu.

Podelio sam podatke po frekvenciji na 10 deleva.

Hteo sam da vidim koje frekvencije su bitnije i koliki bin size je optimalan za svaku frekvenciju

Koristio sam isti model, samo sam input smanjio 10 puta

Bin Size	0	1	2	3	4	5	6	7	8	9
10	30	25	26	26	27	30	26	27	27	28
20	39	29	27	28	29	35	35	31	32	28
40	62	67	64	62	60	67	59	63	63	64
150	61	63	59	58	53	50	60	53	64	50
1200	77	55	55	55	52	46	53	57	64	61

Pretpostavka da su više frekvencije bitnije je tačna.

Takodje sam utvrdio da je bin size od 40 po segmentu skoro dovoljan za predviđanje, sa izuzetkom prvog segmenta gde je optimalan segment veličine 1200

4. Upoređivanje mog rešenja sa onima iz naučne literature

U naučnom radu koji sam pratio uspeli su da dobiju accuracy 88.54% koristeći spektrogram i CNN.

Dobili smo skoro isti accuracy (85%), ali oni su imali više potencijalnih klasa. (9 klasa)

Bitno je napomenuti da su oni trenirali model na nekoliko Gigabajta podataka, za razliku od mojih 400 pesama.

Oni su trenirali na segmentima pesama u trajanju od 30 sekundi, za razliku od mog projekta koji dobija celu pesmu kao input. Oni su odbacili dublje frekvencije, ali sam ja zaključio da to nije poželjno.

Techno i regge pesme model dobro prepoznaje skoro uvek, klasične i metal brka u 30% test primera.

4. Zaključak

Čak i ako sam koristio slične (skoro identične metode) moji rezultati su приметно gori od onih iz literature.

Glavni faktor je verovatno veličina mog trening skupa, ali hteo sam da sam napravim svoj trening skup i da model dobija celu pesmu (ne samo 30 sekundi).

U projektu nisam iskoristio činjenicu da metal pesme obično traju duže od ostalih žanrova, Osećao sam se kao da bi prosleđivanje podatka o dužini pesme bio nizak udarac pa to nisam uradio. Ali model bi definitivno mogao sa sigurnošću da klasifikuje sve pesme iron-maidena.

5. Link rada koji sam pratio

https://dl.wqtxs1xze7.cloudfront.net/59934287/IRJET-V6I517420190704-120568-1u4iafr-libre.pdf?1562308085=&response-content-disposition=inline%3B+filename%3DIRJET_Music_Genre_Classification_using_M.pdf&Expires=1704204956&Signature=VKBhUd2W535brGSW0~wDY6MH0Mx1W0mLIm6NdKZk19IE2bvGBiJY03qG6fX4yT3oDpGi0ab3bbiK-5y0TDB~rzV4zZ0-3rkhSXT6Ynn~Z-PwrSZU2i2TYV5eABDipZxpmk5nycPMd39O4w6VYYNmzl4-7pZgSwUwutACM21Sk3ln6AxPJ1cApQXPZl6iheRNintoGzttHHFjjhb4BSogR58AkTE8Eh1jC1XIUUEHLFTtLpLHQCYY6oKDzTwloO05SiNs30n0izi9~nMXziaGSGNIWU0zvrYT34MiGJ95Rn8E2WnH DcF0pqyhwBYd7wxyvvc9lvYi9vkXI8JM-Po-A_&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA