Mašinsko učenje

Klasifikacija pesama u odgovarajuće žanrove

Nemanja Maček

Sadržaj

Uvodne napomene

Šta je problem sa muzičkim datotekama?

- Do sada smo imali „luksuz“, odnosno nalazili smo se u „zoni udobnosti“ – instance obučavajućih podataka smo lako mogli opisati vektorom obeležja.
- Na primer, u primerima zasnovanim na tekstu, mogli smo da pretvorimo tekst u tzv. „vreću reči“ i generišemo obeležja koja su pogodno „opisivala“ tekst, za npr. klasterovanje, merenje sličnosti itd.
- Postavljamo sledeće pitanje – kako možemo predstavljali pesmu od tri minuta?
- Da li je pristup sličan pristupu tekstualnim podacima, odnosno generisanje „vreće zvučnih bitova“ na osnovu pojedinačnih bitova pesme ispravan?
- Nije, zato što bi postupak bio veoma složen.
- Drugim rečima, potrebano je da pesmu pretvorimo u niz vrednosti koje je dovoljno dobro opisuju, odnosno potrebno je formirati zvučno-zavisna obeležja koja su pogodna za žanrovsku klasifikaciju.

Uvodne napomene

Formulacija problema.

- Pretpostavimo da smo našli gomilu slučajno imenovanih MP3 datoteka na disku za koje pretpostavljamo da sadrže muziku.
- Naš zadatak je da ih klasifikujemo po žanrovima – npr. džez, klasična, kantri, pop, rok, metal, itd.

Uvodne napomene

Muzički žanr.

- Muzički žanrovi su kategorije (kreirane od strane čoveka) koje su nastale kroz složeno međusobno delovanje kulture, umetnosti i marketinga da bi se okarakterisale sličnosti između muzičara ili numera, kao i da bi se organizovale muzičke kolekcije.
- Muzički žanrovi međusobno dele određene karakteristike koje su tipične za instrumentaciju, ritmičku strukturu i tonski sadržaj muzike.
- Problem automatske klasifikacije muzičkih audio zapisa zahteva eksplicitno definisanje žanrova, tj. hijerarhijski set kategorija koji će biti preslikan na muzičku kolekciju.
- U nekim istraživanjima o broju žanrovskih kategorija koje se koriste u muzičkoj industriji pokazano je da nije jasno izgrađena takva hijerarhija žanrova.

Uvodne napomene

Izvođač, album ili numere?

- Jedno osnovno pitanje na koje treba dati odgovor glasi: na koje delove muzike treba primeniti žanr klasifikaciju – na numere, album ili izvođača?
- Ako pretpostavimo da jednu pesmu možemo klasifikovati u samo jedan žanr, to više nije tako jednostavno za jedan album, jer on može biti višežanrovski materijal.
- Isto važi i za izvođača – neki izvođači pokrivaju širok spektar žanrova tokom karijere i nema smisla svrstavati ih u jednu specifičnu klasu.

Uvodne napomene

Neslaganje oko taksonomije.

- Istraživači Pachet and Cazaly u svojim istraživanjima kažu da generalno ne postoji sporazum o taksonomijama žanra koji bi se poštovao u praksi.
- Uzimajući kao primer dobro poznate Web sajtove, kao što su Allmusic (sadrži 531 žanr), Amazon (719 žanrova), Mp3 (430 žanrova), oni su pronalašli samo 70 termina koji su zajednički za sve tri taksonomije.
- Oni takođe primećuju da široko korišćeni termini kao što su rok i pop označavaju različit skup numera i te hijerarhije žanrova su različito organizovane zavisno od konkretne taksonomije (npr. taksonomija Amazona).

Uvodne napomene

Loše definicije žanrova.

- Ako se bliže pogledaju neki specifični i široko korišćeni muzički žanrovi, može se videti koliko je različit kriterijum definisanja specifičnosti žanra.
- Na primer:
  - indijska muzika (engl. *Indian music*) ja geografski definisana,
  - barokna muzika (engl. *Baroque music*) je povezana sa jednim istorijskim razdobljem koje uključuje različite stilove i širok geografski region,
  - post-rok (engl. *Post-rock*) je termin koji je izmislio kritičar Simon Reynolds.
- Ova semantička zabuna između pojedinih taksonomija može da vodi u redundantnost, koja možda neće biti smetnja za ljudski faktor, ali će automatskim sistemima biti veoma otežavajuća okolnost.

Skalabilnost taksonomije žanrova.

- Hijerarhije žanrova takođe trebaju razmatrati mogućnosti dodavanja novih žanrova na račun muzičke evolucije.
- Novi žanrovi se učestalo pojavljuju i oni su:
  - rezultat delimične ili potpune integracije različitih žanrova (npr. *progressive-blackened-death metal*) ili
  - posledica razdvajanja žanrova na podžanrove (npr. *black metal* se deli na *atmospheric black metal*, *symphonic black metal*, *progressive black metal*, itd.).
- Dodavanje novih žanrova i podžanrova u taksonomiju je jednostavno, ali zahteva da adaptaciju automatskog sistema nadgledanog učenja na novonastale promene.

Uvodne napomene

Indeksiranje signala.

- Signal se predstavlja pomoću određenih obeležja, koje oslikavaju određene karakteristike signala bilo u vremenskom, bilo u transformacionom, npr. frekvencijskom, domenu.
- Postupak izdvajanja obeležja iz signala naziva se indeksiranje signala.
- Izdvojena obeležja se koriste za obučavanje klasifikatora, a klasifikacija novih signala se vrši na osnovu njihovih obeležja izdvojenih korišćenjem iste procedure.
- Obeležja koja koristimo mogu se podeliti na:
  - obeležja teksture,
  - obeležja ritma i
  - obeležja tonaliteta.

Obeležja teksture

Šta su obeležja teksture?

- Zvučni signali spadaju u grupu nestacionarnih signala, tj. njihove spektralne karakteristike se menjaju u vremenu.
- Zbog toga se analiziraju na kratkim vremenskim intervalima – tzv. prozorima analize.
- Ukoliko je interval analize dovoljno kratak, može se smatrati da je signal u njemu stacionaran i da su parametri signala konstantni na tom intervalu.
- Za zvučne signale kao što su govor i muzika obično se uzima da je trajanje prozora analize dvadesetak milisekundi.
- Kada se zvučnom signalu intervali sa različitim spektralnim karakteristikama menjaju sa određenom pravilnošću, možemo govoriti o zvučnoj teksturi.

Obeležja teksture

Šta su obeležja teksture?

- Da bi se ova pojava kvantitativno ispitala neophodno je signal posmatrati na većem intervalu koji se naziva prozor teksture.
- Prozor teksture se sastoji od više prozora analize i njegovo trajanje je oko jedne sekunde.
- Istraživanja na ljudskim subjektima su pokazala da je čoveku za prepoznavanje muzičkog žanra potrebno svega tri sekunde muzičkog zapisa.
- Iz ovoga se dolazi do zaključka da čovek za prepoznavanje muzičkog žanra koristi, pored drugih karakteristika audio signala, i upravo opisanu muzičku teksturu.

Obeležja teksture

Spektralni centroid.

- Spektralni centroid se izračunava za svaki prozor analize i predstavlja centar mase amplitudnog spektra tog prozora određenog pomoću kratkotrajne Furijeove transformacije (engl. *Short Time Fourier Transform*, STFS).

$$C_t = \frac{\sum_{k=1}^{N} k \times M_t(k)}{\sum_{k=1}^{N} M_t(k)}$$

- gde indeks t označava prozor analize, a $M_t(k)$ je vrednost amplitudskog spektra prozora t za k-tu diskretnu frekvenciju.

Obeležja teksture

Spektralni centroid.

- U daljem tekstu ćemo pod pojmom diskretna frekvencija smatrati indeks diskretne Furijeove transformacije.
- Veća vrednost ovog obeležja ukazuje na veći udeo visokih frekvencija u spektru signala u prozoru analize.
- Prozori muzičkog signala imaju veću vrednost spektralnog centroida od prozora govornog signala zato što muzički instrumenti proizvode tonove viših frekvencija od ljudskog glasa.

Spektralni *roll-off*.

- Spekralni *roll-off* predstavlja diskretnu frekvenciju R_t ispod koje se nalazi 85% raspodele magnituda signala, tj:

$$\sum_{k=1}^{R_t} M_t(k) \approx 0{,}85 \times \sum_{k=1}^{N} M_t(k)$$

- Vrednost ovog obeležja je veća ukoliko je više energije signala sadržano u visokim frekvencijama.

Obeležja teksture

Spektralni fluks.

- Spektralni fluks odražava promenu spektra između dva susedna prozora analize.

$$F_t = \sum_{k=1}^{N} \left(N_t(k) - N_{t-1}(k)\right)^2$$

- gde je $N_t(k)$ normalizovana magnituda signala u prozoru k, a $N_{t-1}(k)$ normalizovana magnituda signala u prozoru $k-1$.
- Magnituda u svakom prozoru se normalizuju zbirom magnituda signala na svim frekvencijama za dati prozor.
- Ovo obeležje označava dinamiku promene spektra signala.
- Muzički signal se brže menja od govornog i ima veću vrednost ovog obeležja.

Obeležja teksture

Broj prolaza kroz nulu.

- Broj prolaza kroz nulu je obeležje koje se izračunava u vremenskom domenu.
- Njegova vrednost je broj prolazaka signala kroz nulu na datom prozoru.

$$Z_t = \frac{1}{2}\sum_{m=1}^{M}\left|sgn\big(x(m)\big) - sgn\big(x(m-1)\big)\right|$$

- gde je $x(m)$ signal u prozoru t, a M dužina tog prozora.
- Pošto se u govornom signalu smenjuju intervali zvučnog i bezvučnog govora to znači da se smenjuju i intervali sa velikom i malom vrednošću ovog obeležja.
- Broj prolazaka kroz nulu na jednom prozoru je kod muzičkog signala relativno konstantan.

Obeležja teksture

Prozori sa niskom energijom.

- Prozori sa niskom energijom su prozori analize čija je RMS energija manja od prosečne RMS energije u jednom prozoru teksture.
- Ukoliko signal ima veći broj „tihih“ prozora analize, vrednost ovog obeležja će biti veća.
- Veći broj „tihih“ prozora analize karakterističan je za govorni signal.
- Kao obeležje se uzima procentualno učešće ovih prozora u ukupnom broju prozora analize signala.

Obeležja teksture

Mel-skalirani cepstralni koeficijenti.

- Mel-skalirani cepstralni koeficijenti (engl. *Mel Frequency Cepstral Coefficients*, MFCC) su obeležja motivisana ljudskom percepcijom audio signala i često se koriste za modeliranje u sistemima za prepoznavanje govora.
- Da bi se odredili MFCC, signal se propušta kroz filtar-banku čije su centralne frekvencije uniformno raspoređene na logaritamski transformisanoj frekvencijskoj osi.
- Razlog za ovo su eksperimenti na ljudskim subjektima koji su pokazali da uho frekvenciju zvučnih signala opaža na logaritamskoj skali.
- Takođe je pokazano da postoje određeni opsezi frekvencija, kritični opsezi, unutar kojih nije moguće razlikovati frekvencije zvukova.

ISP model MFCC-a.

- ISP (engl. *Intelligent sound implementation*) model realizacije MFCC-a funkcioniše tako što se signal najpre podeli na kratkotrajne prozore dužine N na kojima se izračunava diskretna Furijeova transformacija (DFT):

$$X(k) = \sum_{n=0}^{N-1} w(n)x(n)e^{-j2\pi kn/N}$$

- za $k = 0,1, \dots, N-1$, gde k odgovara frekvenciji $f(k) = k \times f_s/N$ [Hz], pri čemu je f_s frekvencija odabiranja u Hz, a $w(n)$ posmatrani prozor.
- Karakteristike Hammingovog prozora su optimalne pri izračunavanju STFT-a.

ISP model MFCC-a.

- Magnituda $X(k)$ se skalira po frekvenciji i po amplitudi.
- Po frekvenciji se logaritamski skalira korištenjem tzv. Mel filtar-banke $H(k, m)$, a po amplitudi prirodnim logaritmiranjem.

$$H'^{(m)} = \ln\left(\sum_{k=0}^{N-1} |X(k)| \times H(k, m)\right)$$

- za $m = 1,2, \ldots, M$, gdje je M broj filtara u filtar banci i $M \ll N$.

ISP model MFCC-a.

- Mel filtar-banka je kolekcija filtara čije su amplitudne karakteristike trougaonog oblika sa centralnim frekvencijama, date sa:

$$H(k,m) = 0, f(k) < f_c(m-1)$$

$$H(k,m) = \frac{f(k) - f_c(m-1)}{f_c(m) - f_c(m-1)}, f_c(m-1) \leq f(k) < f_c(m)$$

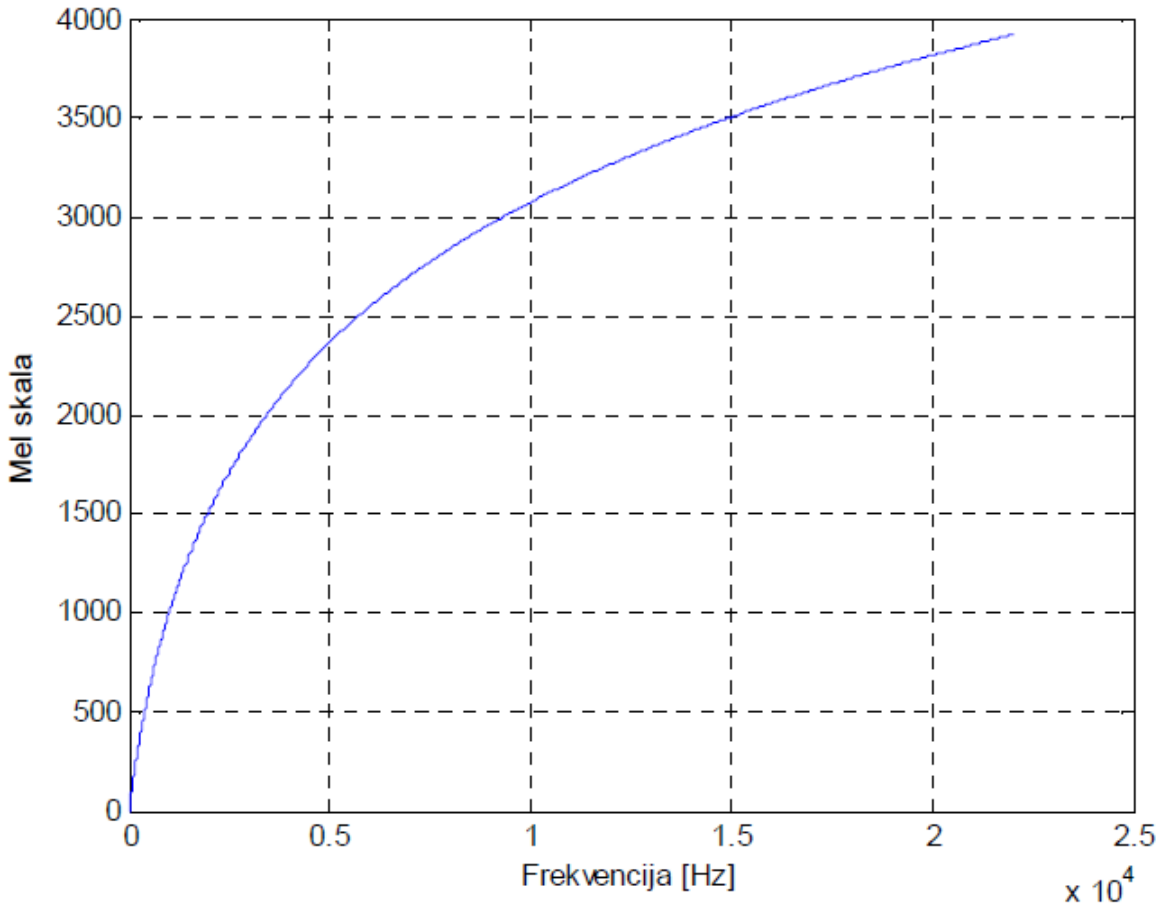
$$H(k,m) = \frac{f(k) - f_c(m+1)}{f_c(m) - f_c(m+1)}, f_c(m) \leq f(k) < f_c(m+1)$$

$$H(k,m) = 0, f(k) \geq f_c(m+1)$$

ISP model MFCC-a.

- Za logaritamsku transformaciju frekvencijske ose koristi se Mel frekvencija koja je sa frekvencijom u hercima povezana jednačinom: $\phi = \log_{10}(1 + f[\mathrm{Hz}]/700)$.
- Frekvencijska rezolucija u Mel skali data je sa: $\Delta\phi = (\phi_{max} - \phi_{min})/(M + 1)$, gde su ϕ_{max} i ϕ_{min} najveća i najmanja rezolucija filtar-banke u Mel skali, izračunate na osnovu f_{max} i f_{min}.
- Ukoliko se, npr, radi klasifikacija žanra u WAV formatu, mogu se postaviti vrednosti: $f_{min} = 0$ i $f_{max} = 22050[\mathrm{Hz}]$.
- Centralne frekvencije u Mel skali date su sa $\phi_c = m \times \Delta\phi$, za $m = 1, \ldots, M$.
- Centralne frekvencije se dobijaju inverznom jednačinom: $f_c(m) = 700 \times \left(10^{\phi_c(m)/2595} - 1\right)$.
- Uvrštavanjem centralnih frekvencija u jednačine na str. 22 dobija se Mel filtar-banka.

ISP model MFCC-a.



ISP model MFCC-a.

- Na kraju, da bi se smanjila dimenzionalnost i korelacija između obeležja, izračunava se diskretna kosinusna transformacija (engl. *Discrete Cosine Transform*, *DCT*) od $X'(m)$:

$$c(l) = \sum_{m=1}^{M} X'(m) \times \cos\left(l \times \frac{\pi}{M}(m - 0{,}5)\right)$$

- za $l = 1, \ldots, M$ gde $c(l)$ predstavlja l-ti MFCC koeficijent.

Obeležja teksture

Srednja vrednost i varijansa obeležja.

- Većina opisanih obeležja su vremenski promenljiva, tj. njihova vrednost se razlikuje u pojedinim prozorima analize u kojima se smatra da je zvučni signal stacionaran.
- Spektralni centroid, spektralni *rolloff*, spektralni fluks, broj prolazaka kroz nulu i MFCC se računaju za svaki prozor analize signala.
- Ova obeležja izračunata za svaki prozor analize mogu poslužiti kao osnova za klasifikator koji bi radio u realnom vremenu.

Obeležja teksture

Srednja vrednost i varijansa obeležja.

- Međutim, ako se projektuje klasifikator koji koristi čitav raspoloživi signal potrebno je izračunati globalna obeležja koja predstavljaju čitav signal.
- Da bi se ovo postiglo koriste se srednja vrednost i varijansa obeležja na prozoru teksture.
- Na ovaj način modeluju su prosečna vrednost obeležja i mjera odstupanja stvarnih vrednosti od prosečne.
- Ovako dobijena obeležja su upotrebljiva na pojedinim prozorima teksture.
- Sa druge strane, procenat prozora sa niskom energijom se izračunava na prozoru teksture i vrednost ovog obeležja se dodaje u vektor obeležja za pojedini prozor teksture.
- Čitav signal se sada opisuje jedinstvenim vektorom obeležja koji predstavlja srednju vrednost opisanih vektora obeležja za prozore teksture.

Obeležja ritma

Algoritam za detekciju tempa.

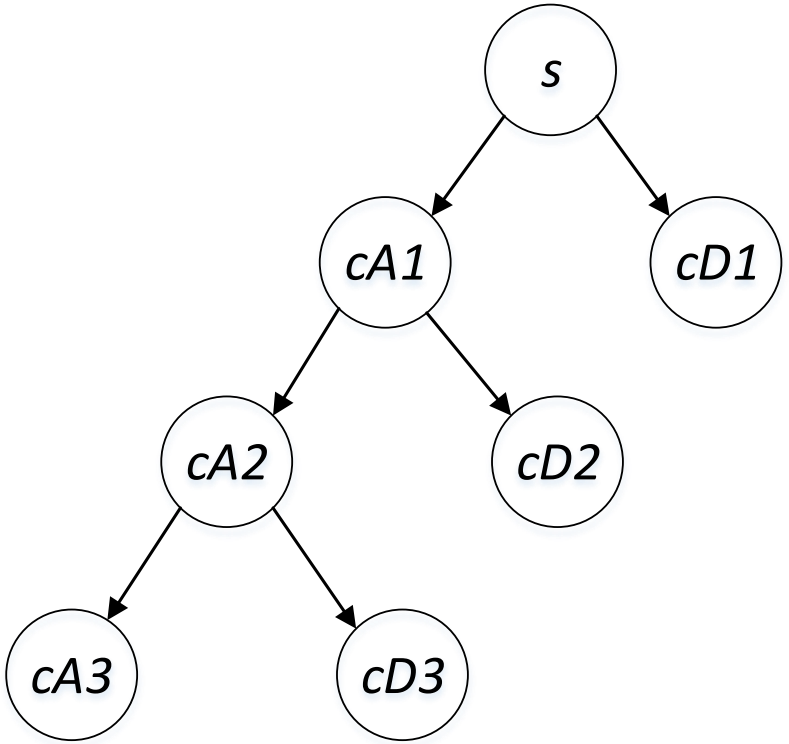
- Jedan od uobičajenih algoritama za automatsku detekciju bita sastoji se od dekompozicije signala pomoću filtar-banke na podopsege (oktave), koja je praćena izdvajanjem envelope signala podopsega i algoritmom koji se koristi za detekciju vremenskih perioda u kojima je anvelopa signala najsličnija samoj sebi.
- Izdvajanje obilježja za reprezentaciju ritmičkog sadržaja iz audio signala zasnovana je na vejvlet transformaciji (engl. *Wavelet Transform*, WT).
- WT je razvijena kao alternativa kratkotrajnoj Furijeovoj transformaciji (STFT) s ciljem da se prevaziđu problemi sa rezolucijom.
- Kod vejvlet transformacije prozor je promenjlive dužine:
  - visoka rezolucija u vremenu a niska u frekvenciji za visoke frekvencije (prozor male dužine),
  - niska rezolucija u vremenu i visoka u frekvenciji za niske frekvencije (prozor velike dužine).

Obeležja ritma

Algoritam za detekciju tempa.

- Diskretna WT je specijalan slučaj WT koja daje kompaktnu reprezentaciju signala u vremenu i frekvenciji i koja može biti uspešno izračunata upotrebom brzog piramidalnog algoritma dekompozicije pomoću filtar-banke.
- Drugim rečima, DWT se može koristiti kao tehnika dekompozicije signala na oktave u frekvencijskom domenu.
- Centralne frekvencije propusnih opsega banke filtara se razlikuju za jednu oktavu (pomnožene faktorom 2).
- U piramidalnom algoritmu dekompozicije, signal je analiziran u različitim frekvencijskim opsezima sa različitom rezolucijom za svaki opseg.
- Ovo je postignuto dekompozicijom signala na grubu aproksimaciju (cAk) i detalje (cDk), zatim se ponovo vrši dekompozicija aproksimacije na novu aproksimaciju i detalje u sledećem nivou.

Algoritam za detekciju tempa.



Algoritam za detekciju tempa.

- Jedan nivo dekompozicije signala obavlja se filtriranjem signala visokopropusnim i niskopropusnim filtrima u vremenskom domenu:

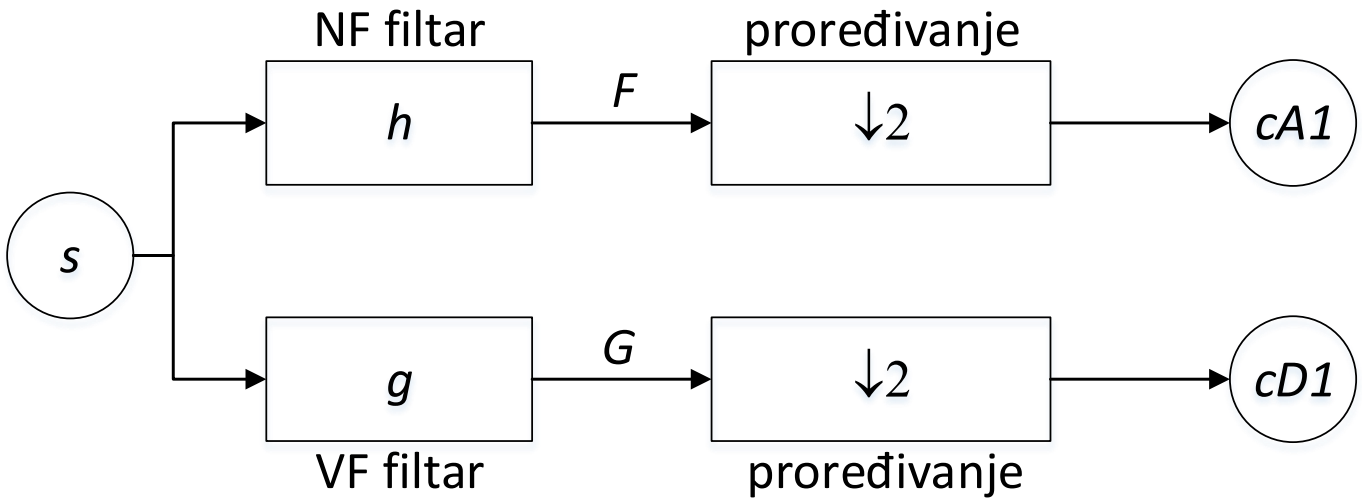
$$y_{high}[k] = \sum_n x(n) \times g(2k - n)$$

$$y_{low}[k] = \sum_n x(n) \times h(2k - n)$$

- gde su $y_{high}[k]$ i $y_{low}[k]$ izlazi visokopropusnog i niskopropusnog filtra, respektivno, nakon proređivanja sa faktorom 2 (v. str. 32).

Algoritam za detekciju tempa.

- cA1 – koeficijent grube aproksimacije, cA2 – koeficijent detalja.



Obeležja ritma

Algoritam za detekciju tempa.

- Algoritam za izdvajanje bita (dobijanje bit histograma) funkcioniše na sledeći način:
  - Nad signalom se prvo izvrši dekompozicija na četiri nivoa (oktave) primjenom DWT-a.
  - Nakon dekompozicije, izdvaja se envelopa signala u vremenskom domenu za svaki opseg posebno.
- Ovo se postiže primenom tehnika:
  - punotalasnog ispravljanja (primenjuje se da se tačnije i lakše izvuče privremena envelopa),
  - niskopropusnog filtriranja (uglađivanje envelope, predstavlja standardnu tehniku izdvajanja envelope sa punotalasnim ispravljanjem),
  - proređivanja u vremenu (koristi se za smanjenje broja odabirka signala radi smanjenja vremena računanje autokorelacije bez efekta na performanse algoritma) i
  - uklanjanja jednosmerne komponente za svaku oktavu (koristi se da bi se signal „centrirao“ na nulu prilikom računanja autokorelacije).

Algoritam za detekciju tempa.

- Nakon toga, envelope svakog opsega su sumirane i izračunata je autokorelacija tako dobijenog signala.
- Autokorelacija je metoda kojom se vrši prepoznavanje periodičnosti (sličnosti) u signalu, tj. tempa (u našem slučaju).
- Dominantni pikovi autokorelacione funkcije odgovaraju različitim periodičnostima envelope signala, tj. bitu koji je sadržan u datom audio signalu.

Obeležja ritma

Algoritam za detekciju tempa.

- Dominantna tri pika (lokalna maksimuma) poboljšane autokorelacione funkcije koji su u rangu za bitsku detekciju, izdvajaju se i dodaju u bit histogram.
- Svaki pik bit histograma odgovara periodu bita u bpm (*beats-per-minute*).
- Na ovaj način, tamo gde je signal sebi najsličniji, pik u bit histogramu će biti najveći.
- Bit histogram daje detaljne informacije o ritmičkom sadržaju, koje mogu biti upotrebljene za klasifikaciju muzičkog žanra.
- Vektor obeležja, baziranih na bit histogramu, izračunava se da reprezentuje ritmički sadržaj i služi za automatsku klasifikaciju muzičkih audio zapisa.

Obeležja tonaliteta

Pitch histogram.

- U sistemima za audio analizu osobine tonaliteta najčešće se izražavaju uz pomoć *Pitch Histograma* (PH).
- PH predstavlja statističku reprezentaciju tonskog sadržaja muzičkog audio zapisa.
- Karakteristike tonaliteta izdvojene iz PH formiraju set obeležja tonaliteta – obeležja izračunata iz PH mogu zajedno sa obeležjima teksture i ritma biti iskorištena za automatsku klasifikaciju žanra muzičkih zapisa.

Obeležja tonaliteta

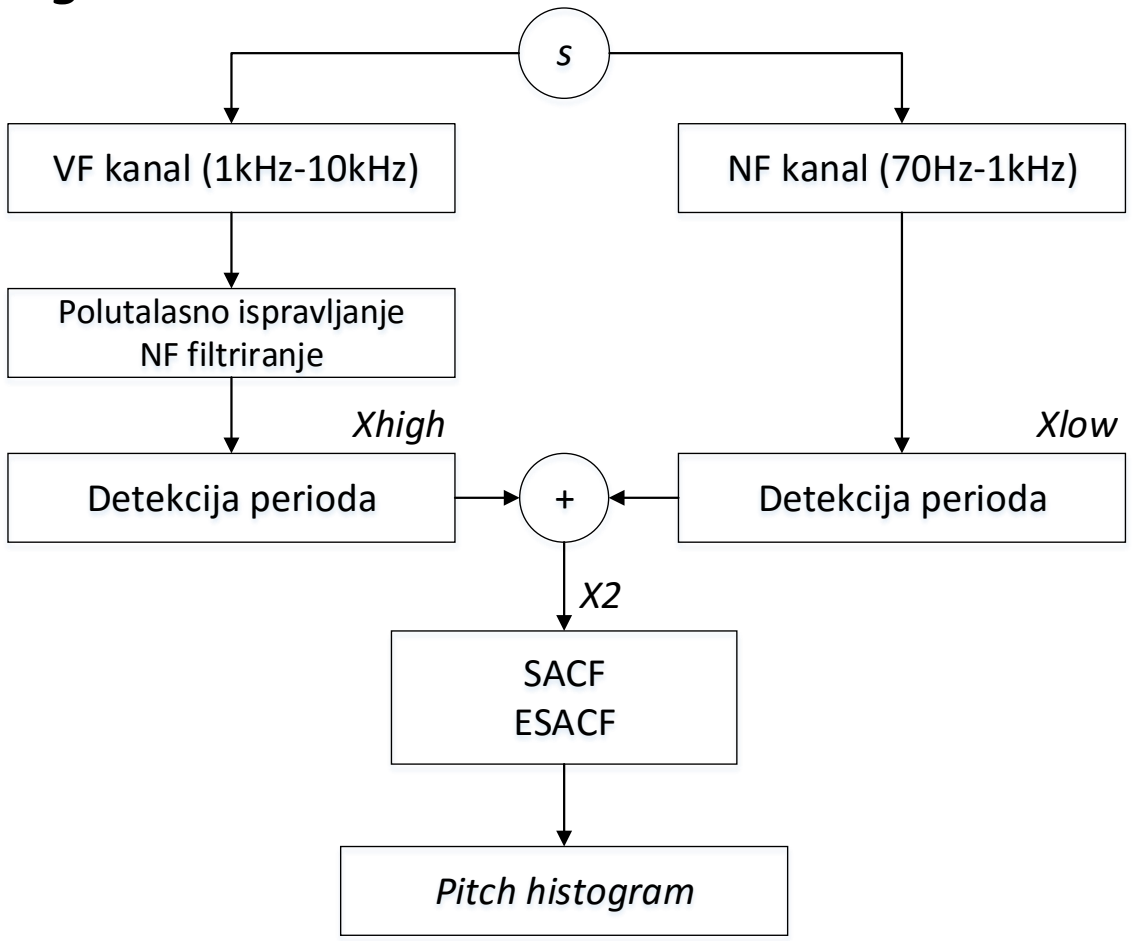
Pitch histogram.

- PH se definiše kao dijagram koji prikazuje zavisnost broja pojavljivanja svake note (tona) u muzičkom audio zapisu od celobrojnih vrednosti (binova) indeksiranih MIDI brojevima (*Musical Instruments Digital Interface*).
- *Pitch Histogram*, u suštini, treba da prikaže tonski sadržaj, odnosno, strukturu muzičkog audio zapisa koja bi trebala da karakteriše određeni žanr.
- Žanrovi sa složenijom tonskom strukturom (kao što su klasika ili džez) imaju raznovrsniji spektar tonova i manje izražene pikove u svojim histogramima nego žanrovi sa „jednostavnijom akordskom progresijom“ kao što su rok, pop ili hiphop.

Multiple Pitch Detection Algorithm.

- Algoritam za izračunavanje PH poznat je pod nazivom *Multiple Pitch Detection Algorithm*.
- Ovaj algoritam zasnovan je na modelu dvokanalne pič (eng. *pitch*) analize.
- Blok dijagram ovog modela prikazan je na str 39.
- Signal se razdvaja na dva kanala, ispod i iznad 1kHz, pomoću filtara propusnika opsega – niskopropusni kanal je dobijen filtrom čiji propusni opseg iznosi od 70Hz do 1KHz, a visokopropusni kanal filtrom čiji je propusni opseg od 1KHz do 10KHz.

***Multiple Pitch Detection Algorithm*.**



Multiple Pitch Detection Algorithm.

- Visokopropusni kanal je još polutalasno ispravljen i „niskopropusno“ filtriran filtrom propusnikom opsega koji se koristio pri odvajanju niskopropusnog kanala.
- Detekcija periodičnosti (engl. *periodicity detection*) bazira se na autokorelacionoj funkciji, tj. izračunava se diskretna Furijeova transformacija (engl. *Discrete Fourier Transform*, DFT), vrši se kompresija magnitude parametrom k, a zatim se primjenjuje inverzna DFT.
- Eksperimentalno je pokazano da optimalna parametra k iznosi $k = 0{,}67$.
- Signal x_2 dat je izrazom:

$$x_2 = IDFT\left(|DFT(x_{low})|^k + \left|DFT\left(x_{high}\right)\right|^k\right)$$

- gde su x_{low} i x_{high} signali pre detekcije periodičnosti u niskopropusnom i visokopropusnom kanalu, respektivno.

Obeležja tonaliteta

Multiple Pitch Detection Algorithm.

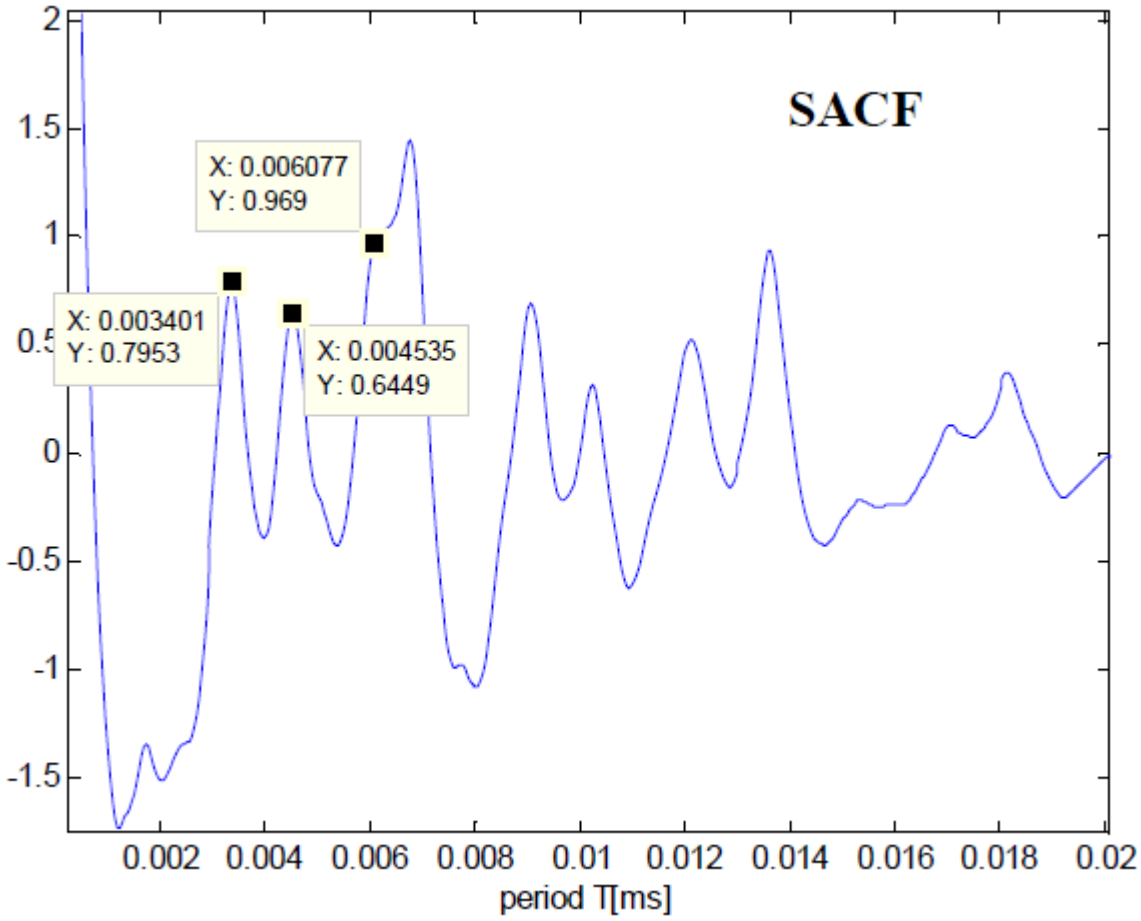
- Pikovi u sumiranoj autokorelacionoj funkciji (*Summary AutoCorrelation Function*, SACF) su relativno dobri indikatori potencijalnih pič perioda u analiziranom signalu.
- Međutim, SACF sadrži redundantne i lažne informacije koje otežavaju utvrđivanje koji pikovi su stvarni pič pikovi.
- Da bi se isključili celobrojni umnošci osnovnog perioda izračunava se poboljšana sumirana autokorelaciona funkcija (ESACF).

Obeležja tonaliteta

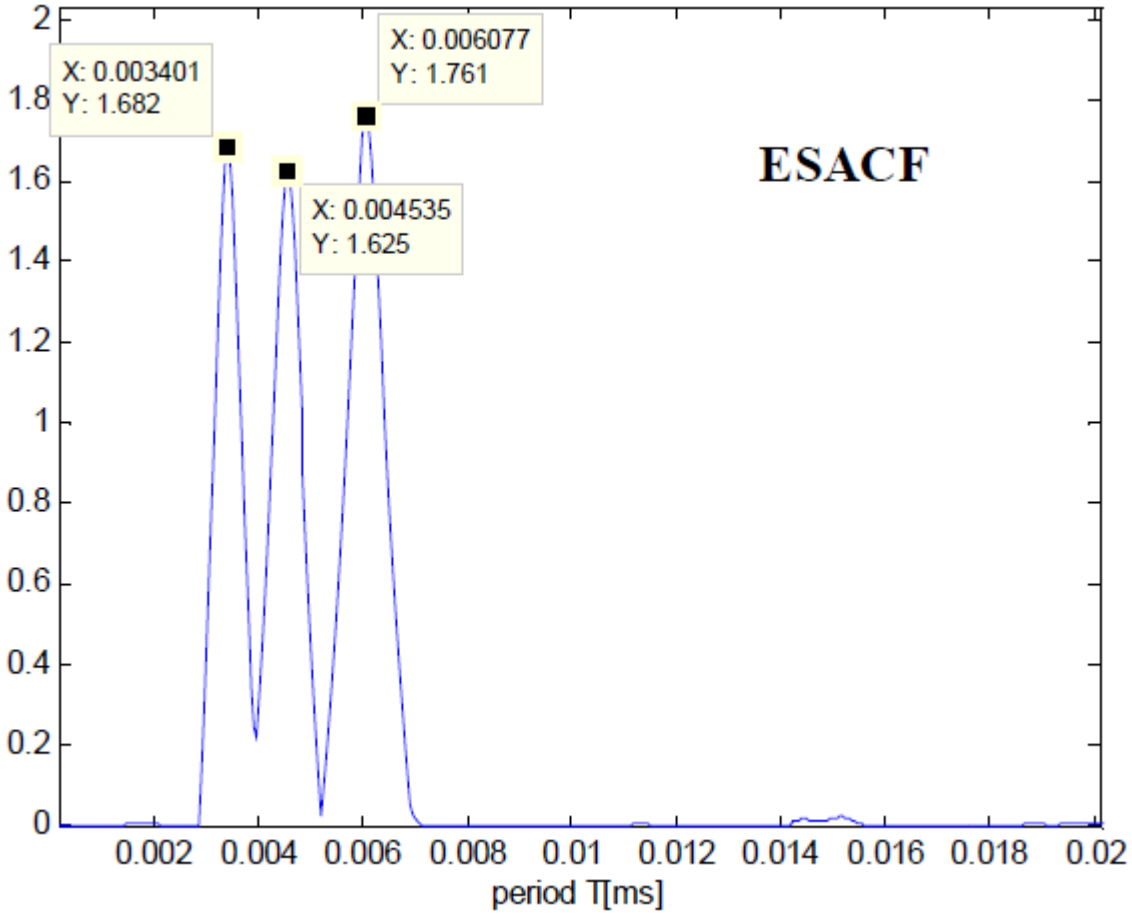
Multiple Pitch Detection Algorithm.

- Primer SACF i ESACF tri tona gitare (E3, A3 i D4) sa fundamentalnim frekvencijama na 165Hz (T=6.07ms), 220Hz (T=4.54ms) i 294Hz (T=3.4ms).

Multiple Pitch Detection Algorithm.



Multiple Pitch Detection Algorithm.



Multiple Pitch Detection Algorithm.

- Kada je dobijena ESACF uzimaju se tri dominantna pika iz svakog prozora analize i stavljaju u histogram.
- Tamo gde se pikovi budu najviše poklapali, amplituda u histogramu će biti najveća.
- Frekvencije koje odgovaraju svakom piku histograma su konvertovane u muzički ton, tako što svaki bin PH odgovara muzičkoj noti odgovarajuće frekencije (na primer, A4=440Hz).
- Muzičke note su definisane MIDI notnim sistemom, a konverzija frekvencije u MIDI notni broj izvršena je jednačinom:

$$n = 12 \log_2 \left(\frac{f}{440} \right) + 69$$

- gde je f frekvencija a n histogram bin (MIDI notni broj).

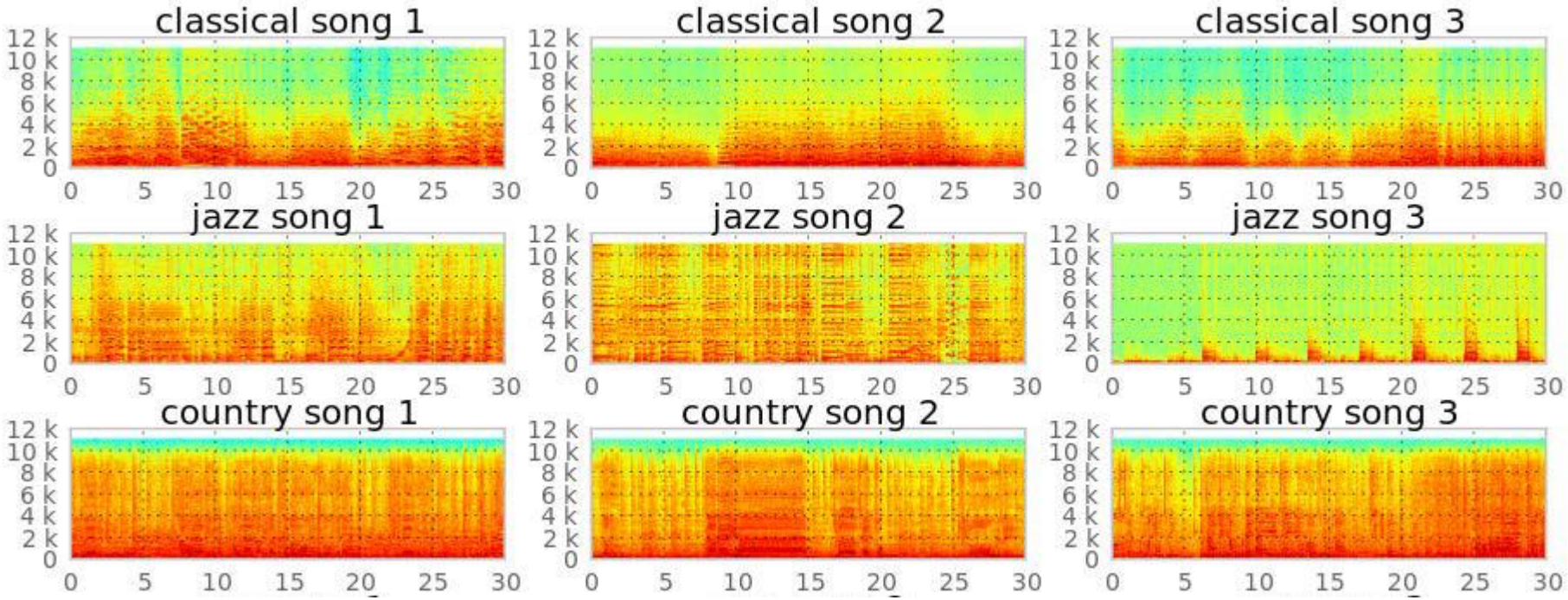
Klasifikacija GTZAN skupa podataka

O skupu podataka.

- GTZAN se često koristi kao skup podataka za klasifikaciju muzičkih žanrova.
- Sadrži 10 žanrova, sa po 100 pesama, pri čemu je za svaku pesmu dato prvih 30 sekundi u formatu 22050 Hz, mono.
- U eksperimentima je korišćeno 6 žanrova: klasična muzika, džez, kantri, pop, rok i metal.
- Skup podataka je dostupan na adresi: http://opihi.cs.uvic.ca/sound/genres.tar.gz.

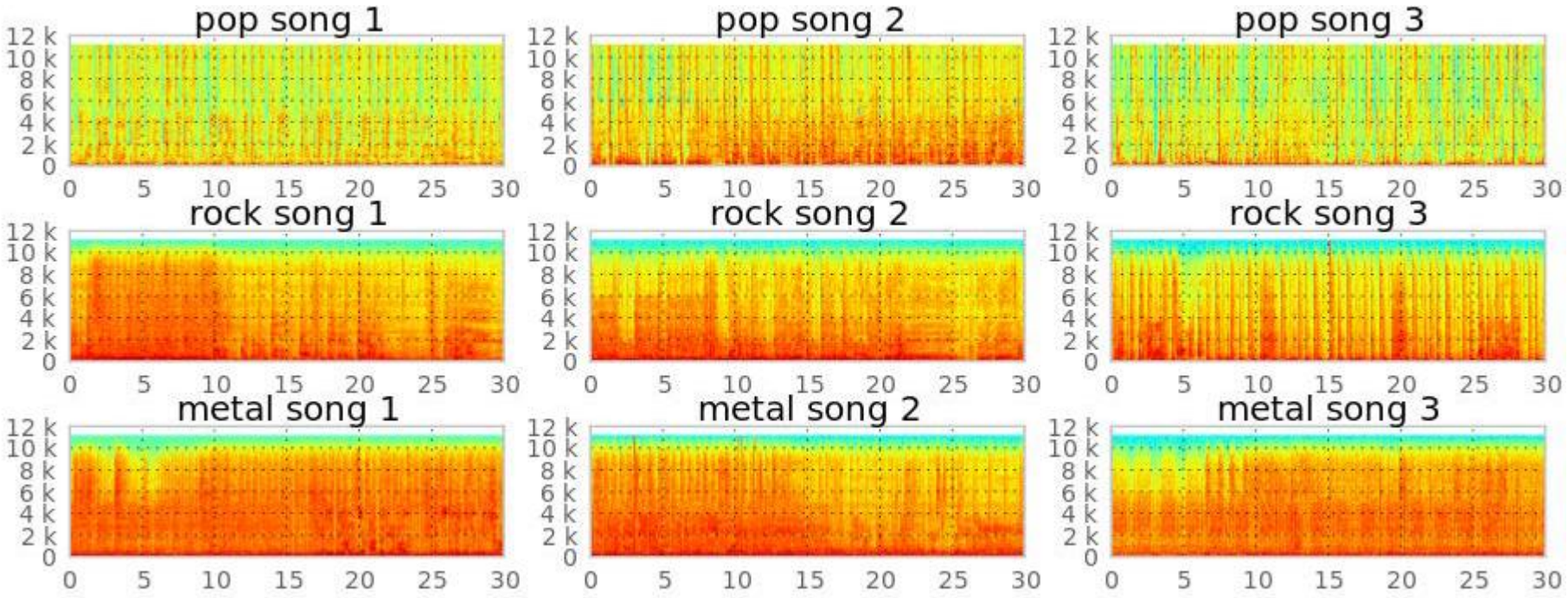
Klasifikacija GTZAN skupa podataka

Spektrogram uzoraka.



Klasifikacija GTZAN skupa podataka

Spektrogram uzoraka.



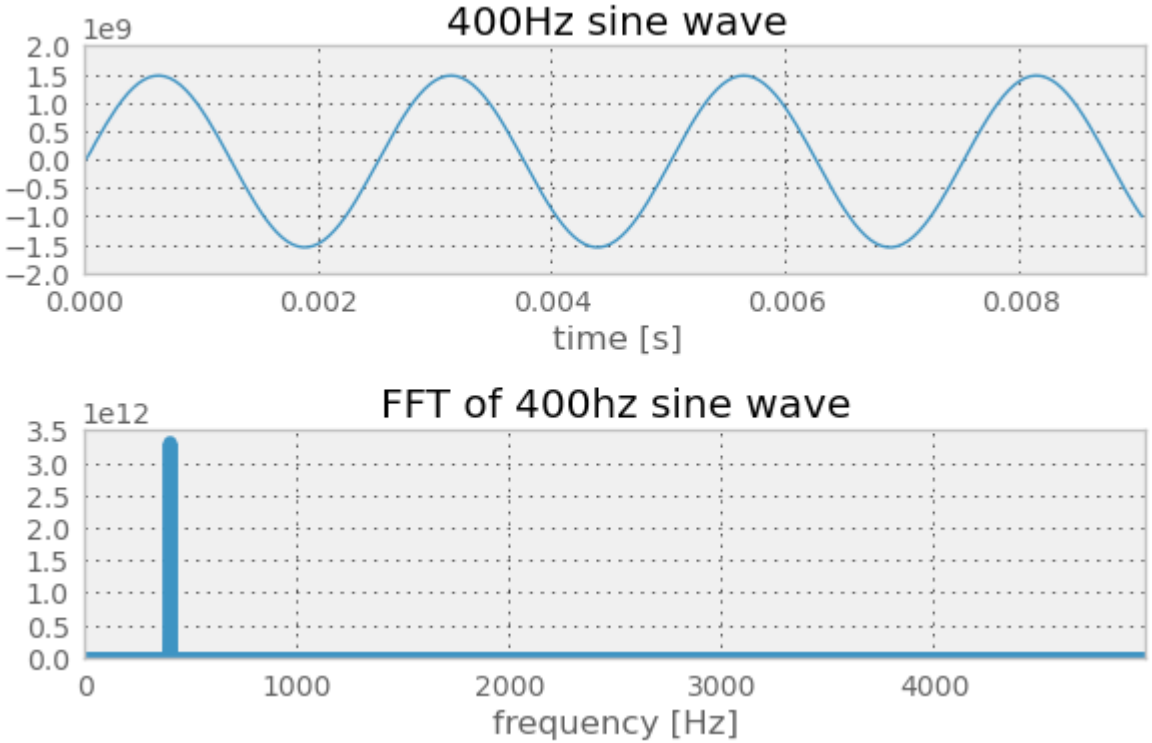
Klasifikacija GTZAN skupa podataka

Šta zaključujemo na osnovu spektrograma?

- Odmah se vidi razlika u spektru između, na primer, metal i klasične numere – dok metal numere imaju veliki intenzitet tokom većeg dela frekvencijskog spektra (energične su), spektar klasičnih pesama je više raznolik tokom vremena.

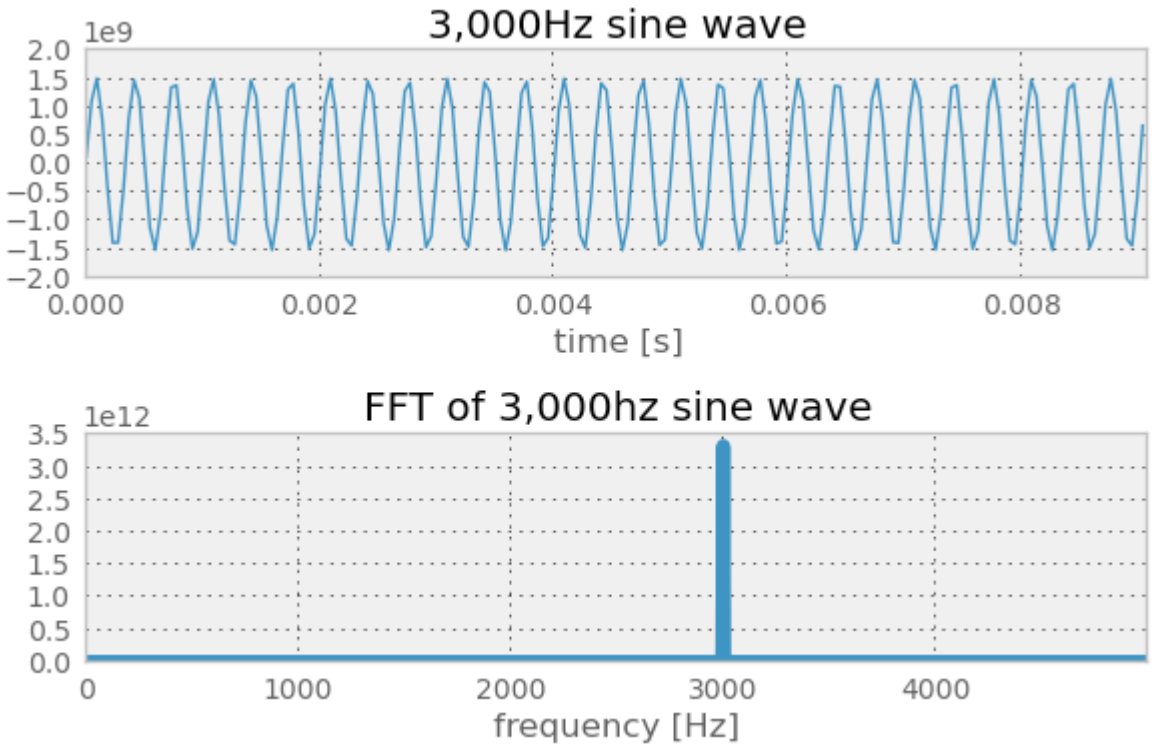
Klasifikacija GTZAN skupa podataka

Izdvajanje obeležja pomoću FFT.



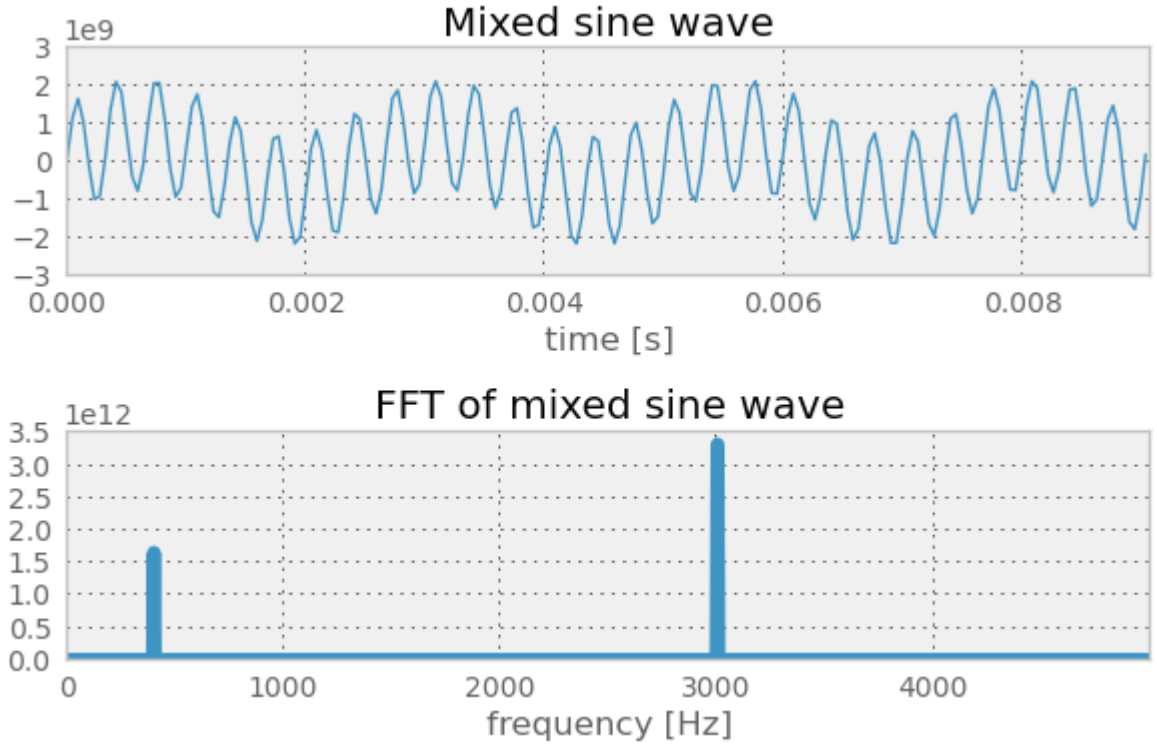
Klasifikacija GTZAN skupa podataka

Izdvajanje obeležja pomoću FFT.



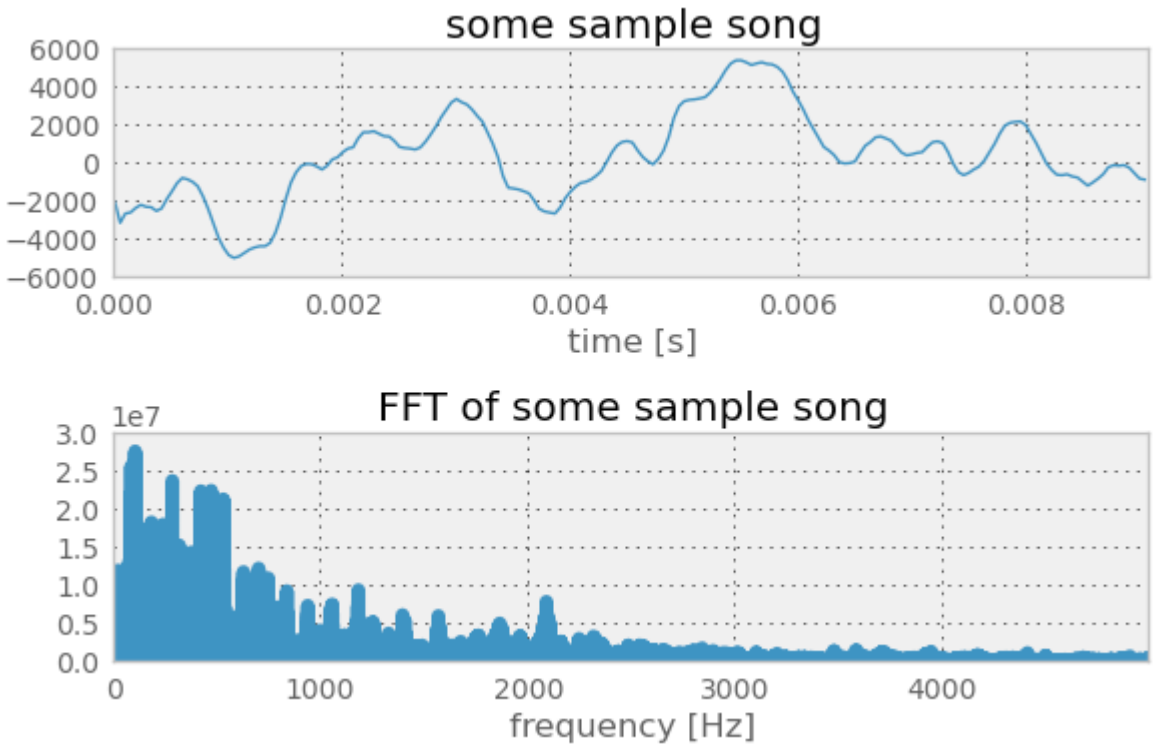
Klasifikacija GTZAN skupa podataka

Izdvajanje obeležja pomoću FFT.



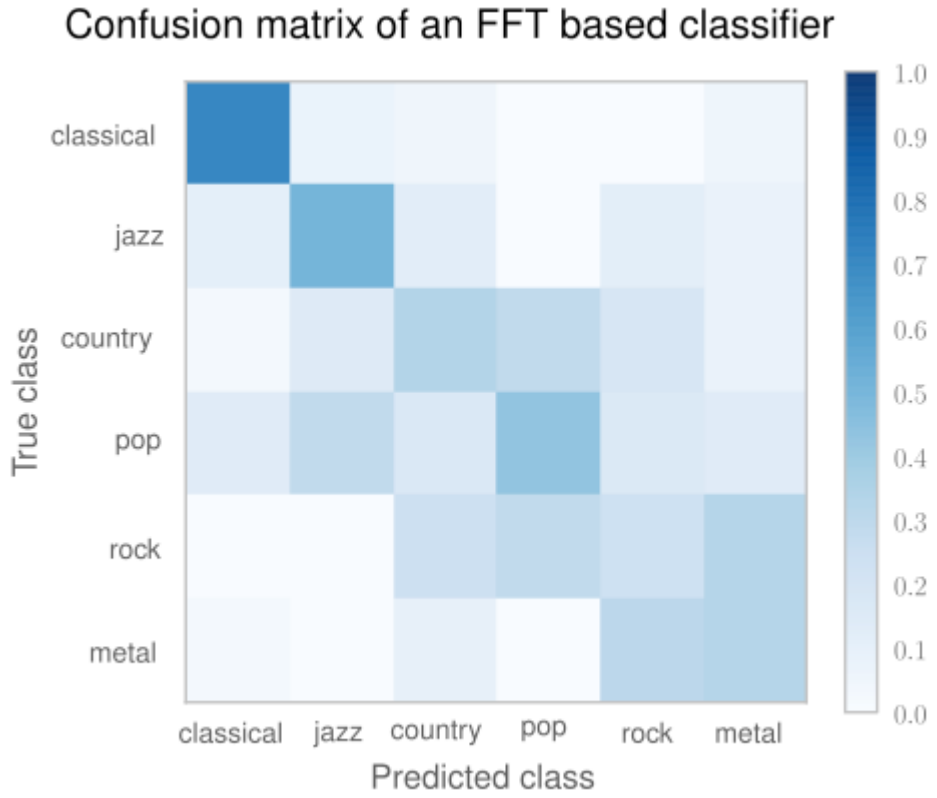
Klasifikacija GTZAN skupa podataka

Izdvajanje obeležja pomoću FFT (realniji slučaj).



Klasifikacija GTZAN skupa podataka

Matrica konfuzije (za logističku regresiju i FFT).



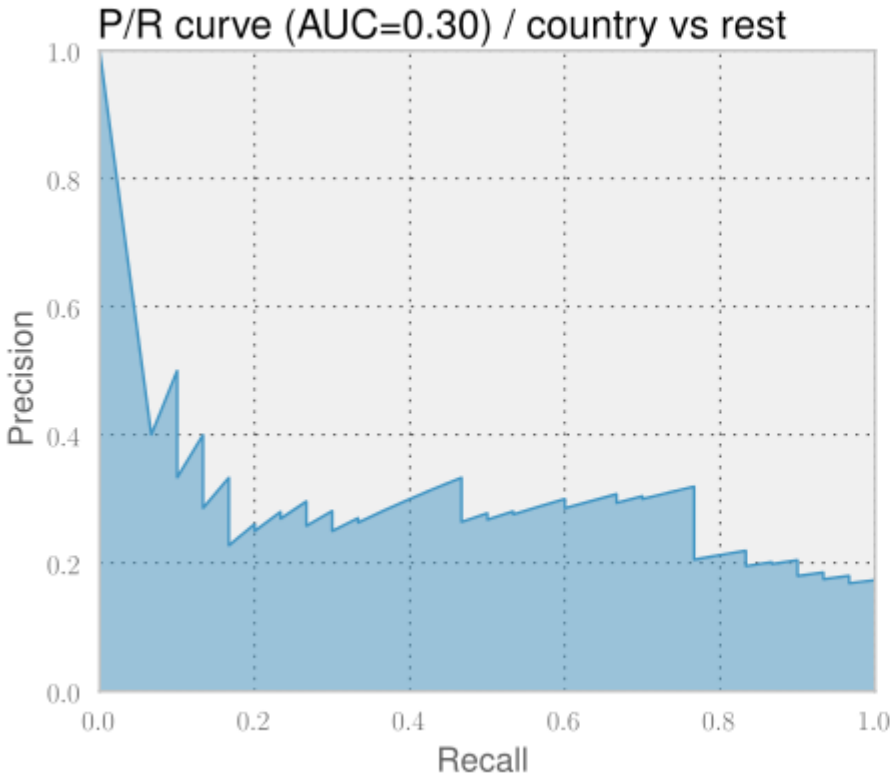
Klasifikacija GTZAN skupa podataka

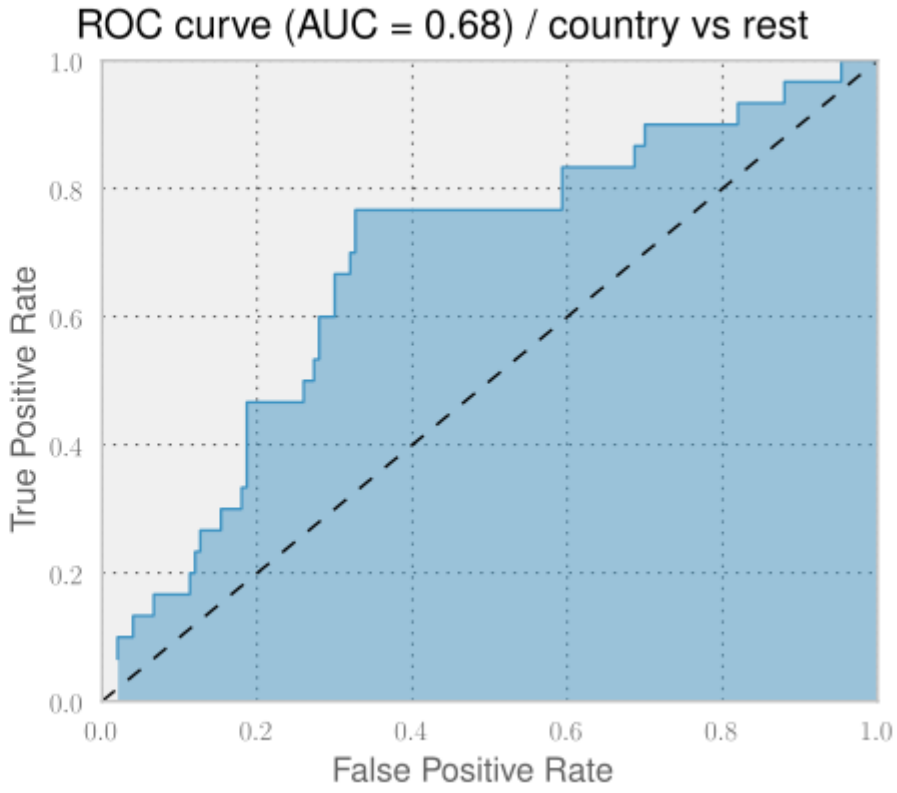
Podsetnik.

Kriva	x-osa	y-osa
Precission / Recall	Recall = TP / (TP + FN)	Precission = TP / (TP+FP)
ROC	FPR = FP / (FP + TN)	TPR = TP / (TP + FN)

Klasifikacija GTZAN skupa podataka

ROC i P/R krive (za logističku regresiju i FFT, kantri protiv ostalih).





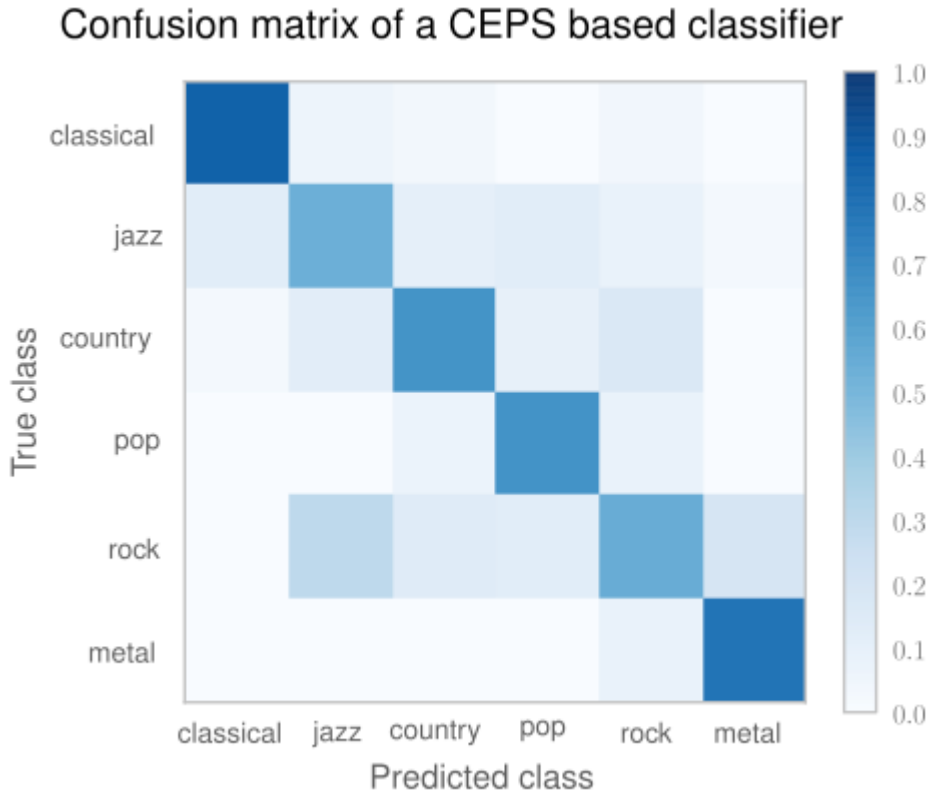
Klasifikacija GTZAN skupa podataka

Šta dalje?

- Ovaj klasifikator jedino dobro diskriminiše klasičnu muziku od ostatka numera.
- Drugim rečima, neupotrebljiv je.
- To znači da izdvajanje obeležja pomoću FFT nije dovoljno dobro rešenje.
- Ubacujemo u igru Mel-skalirane cepstralne koeficijente (MFCC).

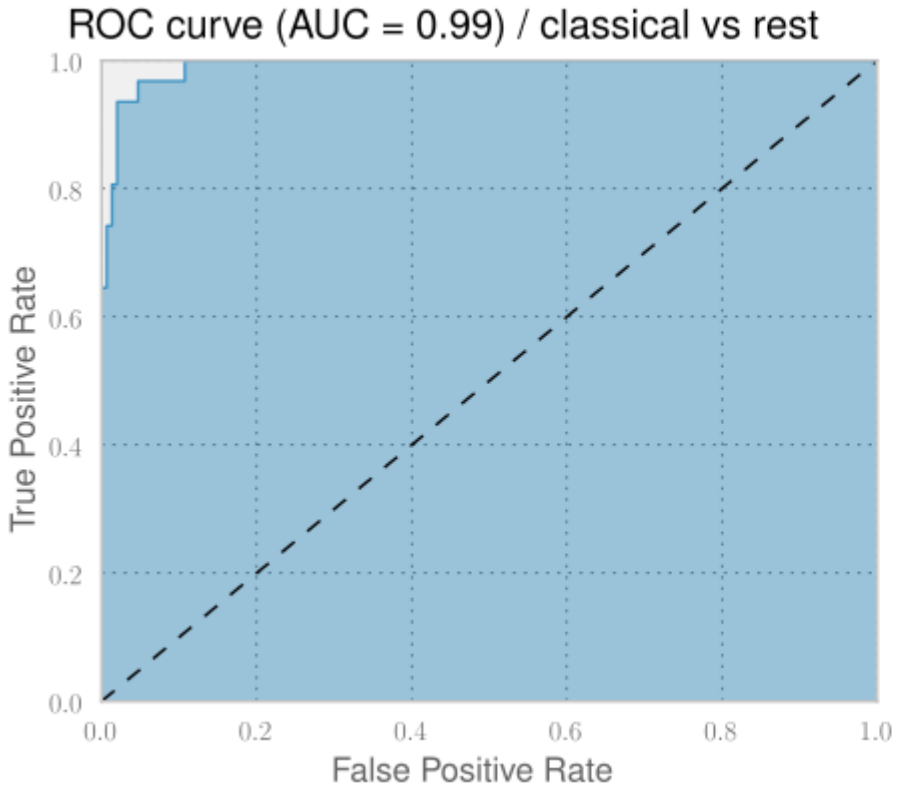
Klasifikacija GTZAN skupa podataka

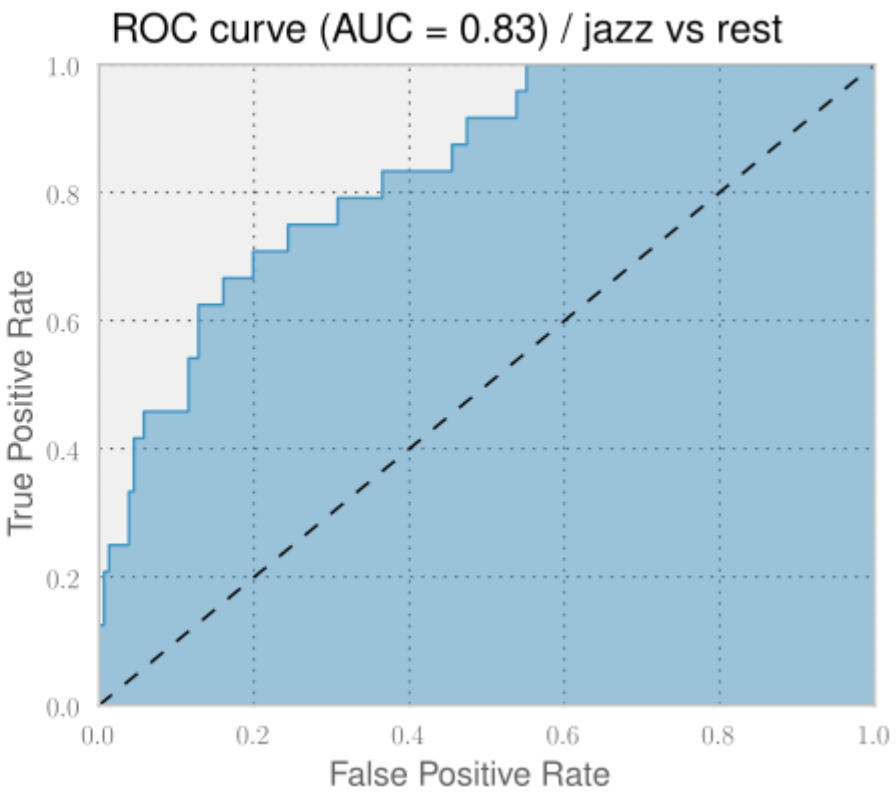
Matrica konfuzije (za MFCC).



Klasifikacija GTZAN skupa podataka

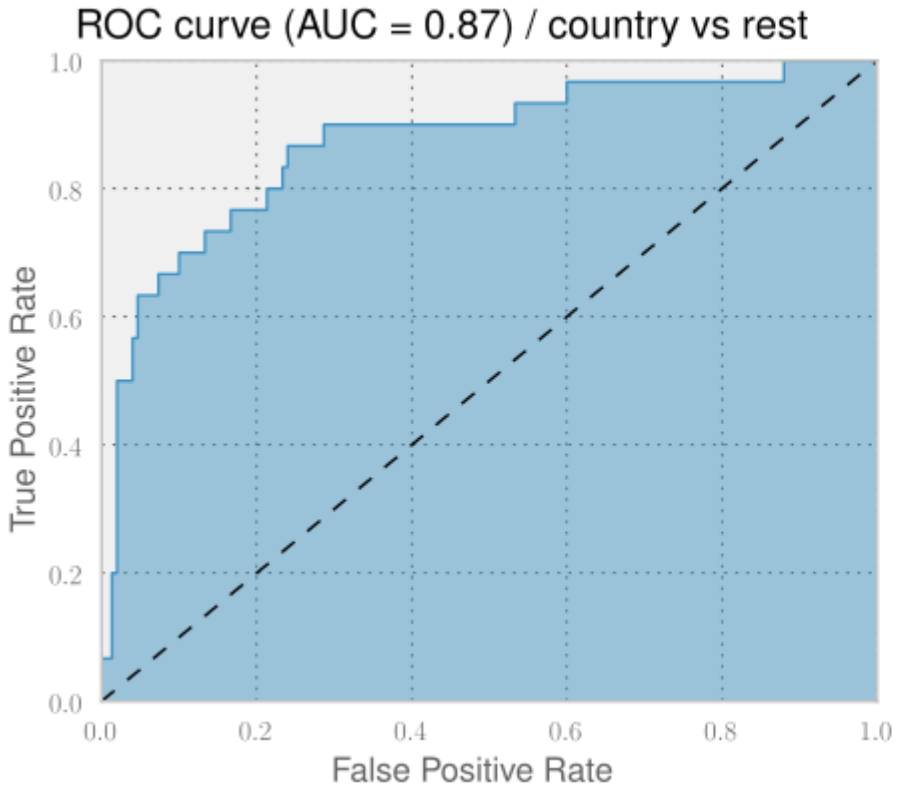
ROC krive (za MFCC).

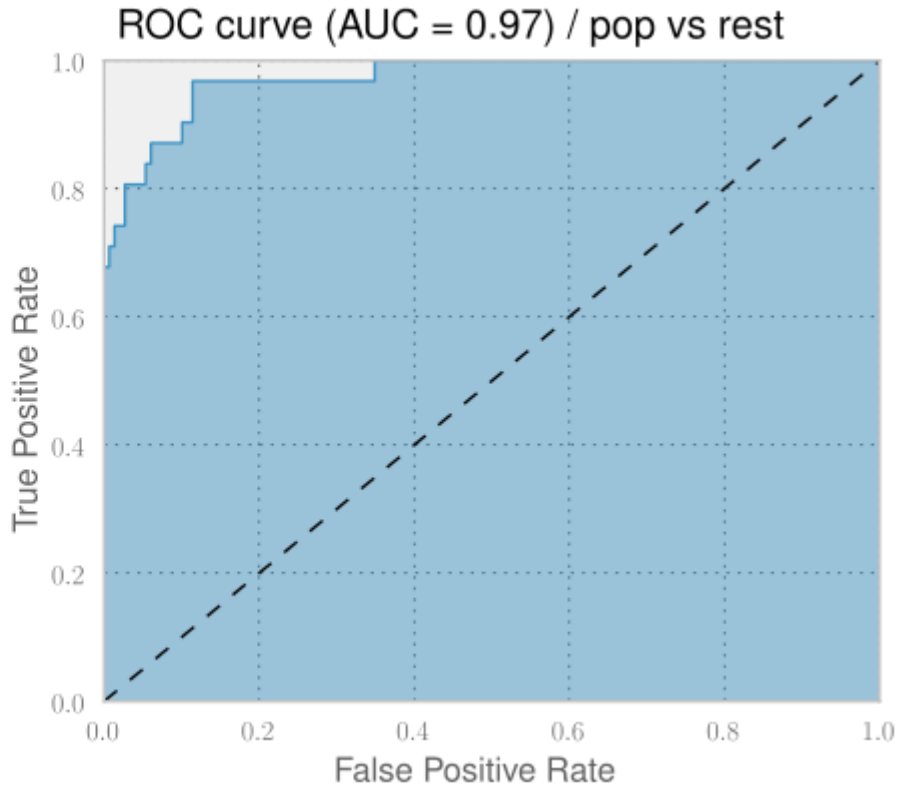




Klasifikacija GTZAN skupa podataka

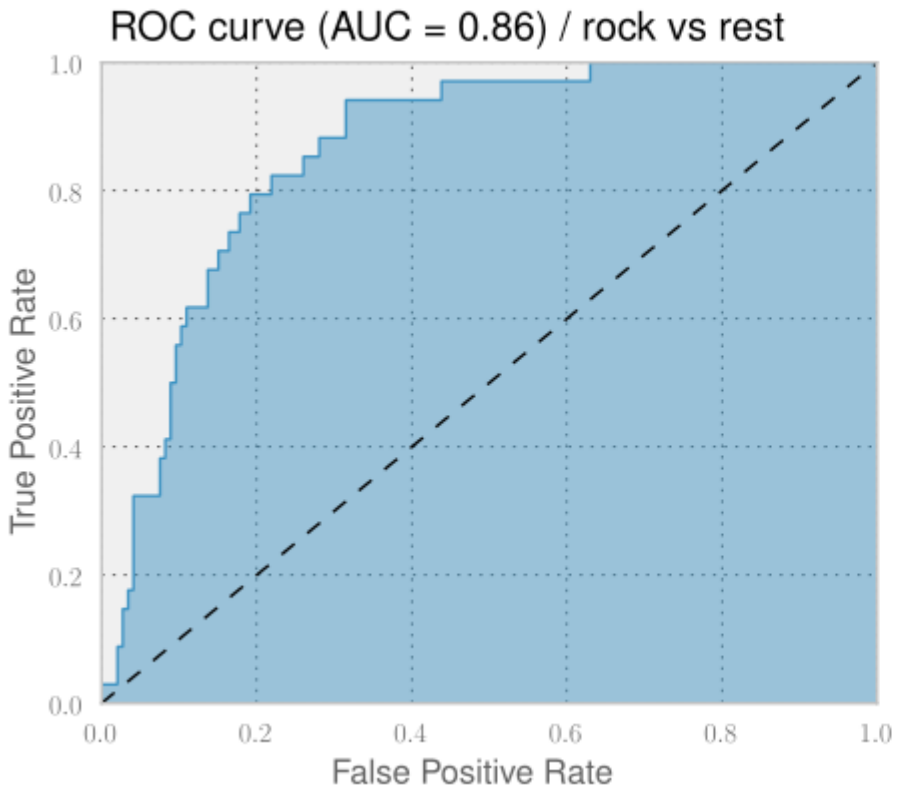
ROC krive (za MFCC).

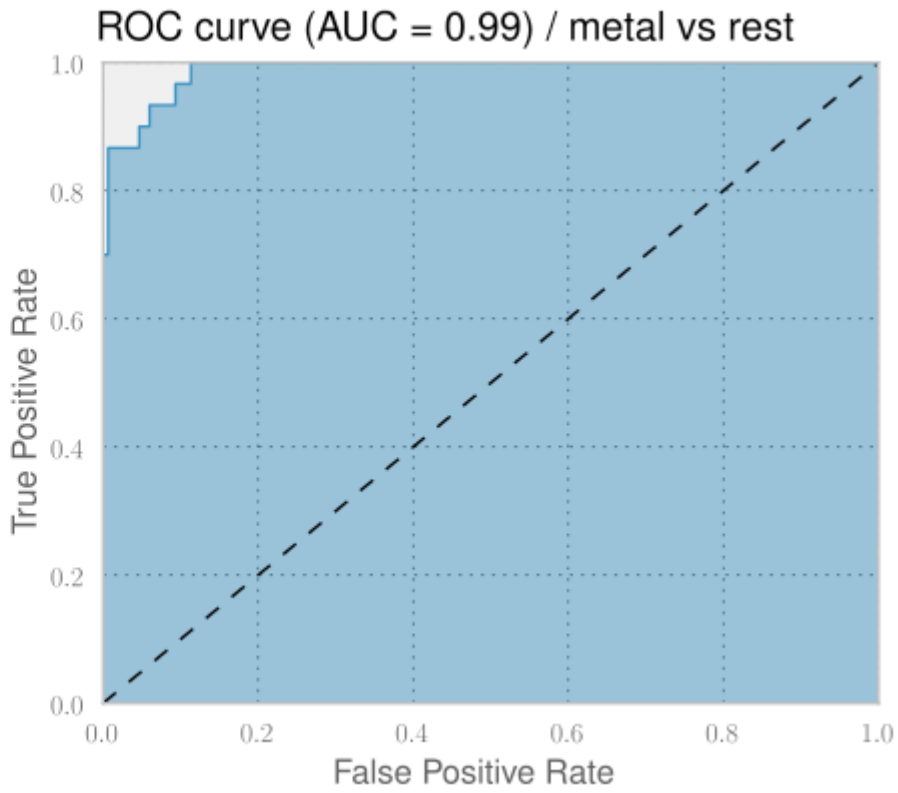




Klasifikacija GTZAN skupa podataka

ROC krive (za MFCC).





Klasifikacija GTZAN skupa podataka

Šta smo postigli?

- Performanse klasifikacije za sve žanrove su poboljšane.
- AUC za klasične i metal numere su ≈1,0 AUC.
- Matrica konfuzije izgleda mnogo bolje.
- Jasno možemo videti dijagonalu koja ukazuje da klasifikator uspeva u većini slučajeva ispravno da klasifikuje žanrove. Ovaj klasifikator je prilično upotrebljiv da reši naš početni zadatak.

Zaključne napomene

- Na ovom izlaganju, potpuno smo izašli „zone udobnosti“.
- Pod pretpostavkom nedostatka dubokog razumevanja akustike i muzičke teorije, koristili smo funkcije koje smo razumeli dovoljno samo da znamo kako i gde ih trebamo da ih stavimo u našu postavu klasifikatora – pri tome, FFT se nije pokazao kao dobar ekstraktor obeležja, dok je MFCC u tome uspeo (razlika između njih je da se u drugom slučaju oslanjali na obeležja koja su kreirali eksperti).
- To je u redu – ako nas zanima rezultat, a ne teorija, ponekad jednostavno moramo da koristimo prečice – samo moramo da budemo sigurni da koristimo prečice na koje su nas odveli stručnjaci u određenim problemskim domenima.

Literatura

- Beleške pripremljene prema knjizi – Luis Pedro Coelho, Willi Richert (2015): „Building Machine Learning Systems with Python, Second Edition“. Packt Publishing.
- Dodatna literatura korišćena prilikom prirpreme beleški: http://dsp.etfbl.net/students/maric.pdf

Hvala na pažnji

Pitanja su dobrodošla.