

Izveštaj

Klasifikacija audio snimaka ponašanja beba

Nikolina Radulović, IN11/2018, radulovicnikolina@gmail.com

Lana Kovačević, IN38/2018, kovaceviclana2501@gmail.com

I UVOD

Audio detekcija i klasifikacija je aktuelna i značajna oblast istraživanja i danas nalazi široku primenu u okviru različitih praktičnih scenarija. Porast interesovanja za probleme ove vrste javlja se zahvaljujući virtuelnim asistentima kao što su Alexa, Siri i Google Home i sličnim tehnologijama zasnovanim na automatizaciji audio obrade. Za potrebe ovog projekta raspoznavali smo da li vrednosti relevantnih obeležja izvučenih iz audio snimaka ukazuju na bebin plač, smeh ili neki neutralan zvuk. Takav algoritam koji je u stanju da automatski detektuje plač novorođenčeta danas nalazi primenu u bezbednosnim uređajima npr. bebi monitorima i deo je sistema za otkrivanje dece zaboravljene u vozilima. Identifikacija praćena kalsifikacijom plaća novorođenčeta može biti od koristi i u medicinske svrhe - otkrivanje patologija na osnovu auditivnih svojstava signala plaća ili procena neurološkog stanja odojčadi na osnovu razlika u plaču između pravovremeno i prevremeno rođenih beba.

II ANALIZA PODATAKA

U okviru tri fajla opisani su audio snimci bebinog plaća, smeha ili neutralnih zvukova, tako da se uzorci iz iste klase nalaze u zasebnom fajlu. Svaki uzorak reprezentovan je obeležjima izvučenim pomoću Praat programskog alata za analizu audio snimaka: osnovna frekvencija, prvi i drugi formant i 12 mfcc (*Mel-frekvencijski kepralni koeficijenti*). Navedena obeležja izvučena su na prozorima od 25ms sa preklapanjem od 10ms, nakon čega su na nivou celog uzorka izračunate njihove statističke vrednosti: srednja vrednost, standardna devijacija, minimum, medijan i maksimum. Na ovaj način dobijeno je 75 kolona koje opisuju 64 uzorka. Nedostajućih vrednosti obeležja nema, upravo zbog navedene procedure formiranja baze podataka. Brojnost svake klase ne potpuno idealno balansiranog skupa podataka data je u Tabeli 1.

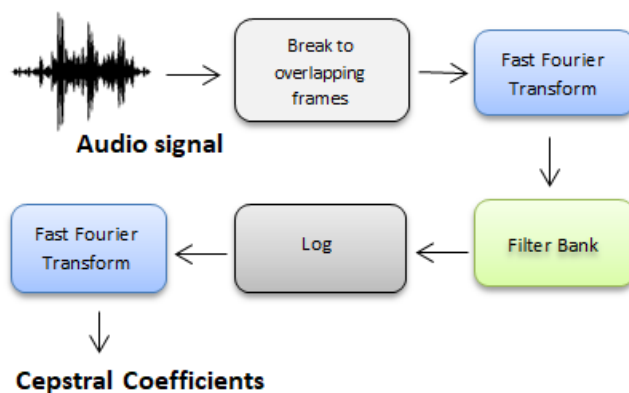
Tabela 1: Brojnost uzoraka svake klase

| Klasa | Broj uzoraka |
|-----------|--------------|
| plač | 21 |
| smeh | 18 |
| neutralan | 25 |

Nad prozorima izvučena navedena obeležja iz oblasti spektralne analize u stanju su da dovoljno detaljno opišu uzorke kako bi algoritmi mašinskog učenja bili u stanju da uoče pravilnosti na osnovu kojih će odrediti klasnu pripadnost novih uzoraka.

Formanti, bilo da se posmatraju kao akustične rezonancije vokalnog trakta, ili kao lokalni maksimumi spektra, predstavljaju karakteristične frekvencijske pojaseve u okviru signala.

Kepstralni koeficijenti Mel-frekvencije (MFCC) zajedno čine reprezentaciju kratkoročnog spektra snage zvuka i na Slici 1 prikazan je pozadinski proces njihovog ekstrahovanja. U bazi je dato prvih 12 mfcc, što je optimalno za manji broj uzoraka i rezultuje manje kompleksnim modelom.



Slika 1. Proces ekstrahovanja MFCC iz audio signala

III ALGORITMI ZA REŠAVANJE PROBLEMA KLASIFIKACIJE

Problem svrstavanja audio snimaka u jednu od grupa ponašanja beba je problem klasifikacije. Problem se rešavao korišćenjem tri veoma zastupljena algoritma, a to su:

- KNN algoritam,
- Logistička regresija,
- Neuronske mreže.

A. KNN ALGORITAM

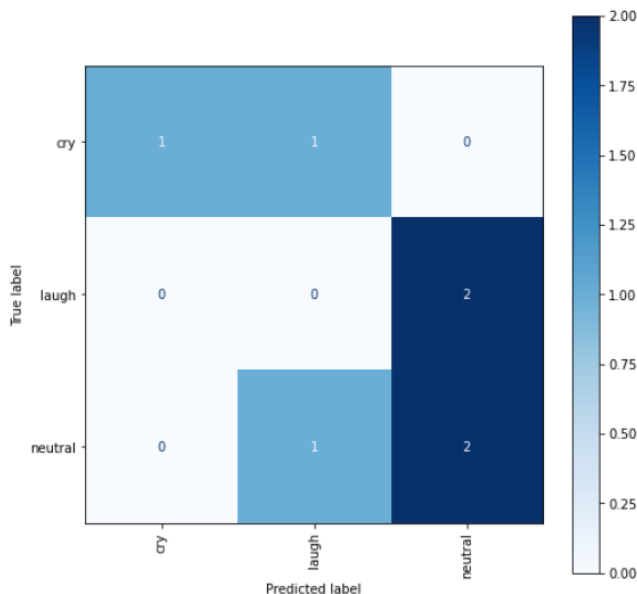
K - Nearest Neighbour klasifikator klasifikuje nepoznati uzorak na osnovu klasne pripadnosti susednih uzoraka iz skupa za obuku, tako da se uzorak dodeljuje klasi koja je u većini među *k* najbližih uzoraka. Prilikom instanciranja kNN klasifikatora potrebno je zadati parametre - broj suseda i metriku, za čiji odabir je primenjena metoda unakrsne validacije.

Početni skup je prvenstveno podeljen na trening i test skup. Primenom unakrsne validacije na trening skupu i analizom mere tačnosti klasifikacije utvrđeno je da je optimalan broj suseda koji se uzima u obzir pri odlučivanju 1, a metrika za utvrđivanje rastojanja između dva uzorka euklidska, računata po formuli:

$$d_{xy} = (\sum_{i=1}^D |x_i - y_i|^2)^{\frac{1}{2}}$$

Prosečna tačnost po klasi modela je 76%, najveća tačnost odgovara klasi *cry* - 80%, dok je najmanja tačnost klase *neutral* - 70%. Samim tim, vidimo da je najviše ispravno klasifikovanih uzoraka u celoj populaciji upravo iz klase *cry*.

Razmatranjem mere osetljivosti uvidjamo da u proseku po klasi iznosi 63%, gde se uočava niska vrednost osetljivosti najmalobrojnije klase *laugh* - 56%, što objašnjava performanse algoritma na test skupu gde su oba uzorka iz klase *laugh* pogrešno klasifikovana i dodeljena najmnogobrojnijoj klasi *neutral*.

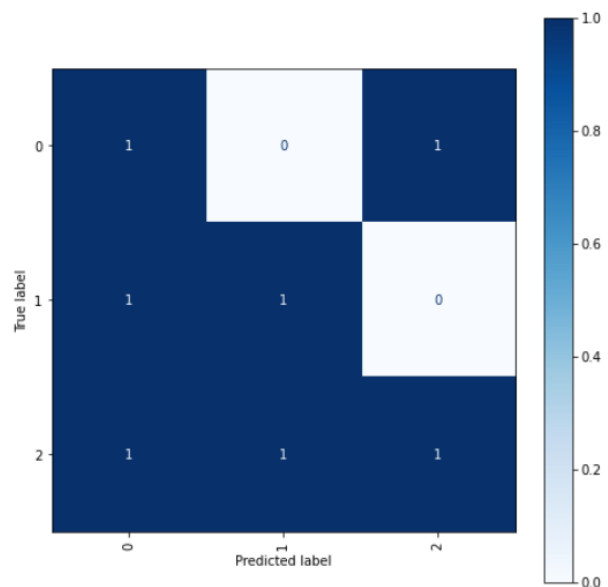


Slika 2. Rezultati kNN klasifikatora na test skupu

B. LOGISTIČKA REGRESIJA

Logistička regresija je nadgledani algoritam učenja koji za dati uzorak na ulazu predviđa verovatnoću pripadnosti klasama. Unakrsnom validacijom određeni su optimalni parametri tj vrednost broja iteracija i odgovarajući algoritam za rešavanje problema gde je ustanovljeno da je 4000 iteracija potrebno da algoritam konvergira ka rešenju ako je solver *sag*.

Najveća tačnost odgovara klasi *cry*, 88% a isto važi i za osetljivost, 73%. Osetljivost klase *laugh*, za razliku od uspešnosti preostala dva algoritma nije drastično manja, već sasvim prosečna, ali sam prosek osetljivosti po klasama je nešto niži, 68%, pa se može očekivati da će model gotovo jednako grešiti pri klasifikaciji uzoraka svake od klasa.



Slika 3. Rezultati logističke regresije na test skupu

C. NEURONSKE MREŽE

Neuronske mreže se danas u praksi najčešće primenjuju za rešavanje ovakvih problema. U ovom slučaju, kako bi izvršili klasifikaciju uzoraka, implementirali smo nerekurzivne višeslojne neuronske mreže uz pomoć klase *MLPClassifier* namenjene za primenu u klasifikacionim problemima. Navedena klasa omogućava podešavanje većeg broja parametara - broj slojeva, broj neurona po sloju, aktivaciona funkcija pomoću koje mreža samostalno pronalazi optimalne težine, solver tj. algoritam za ažuriranje parametara mreže, broj uzoraka jednog *batch*-a i dr. Eksperimentalnim putem određeno je da optimalne

rezultate daje mreža sa 4 sloja i 75 neurona po sloju, koja tačno predviđa 62% uzoraka.

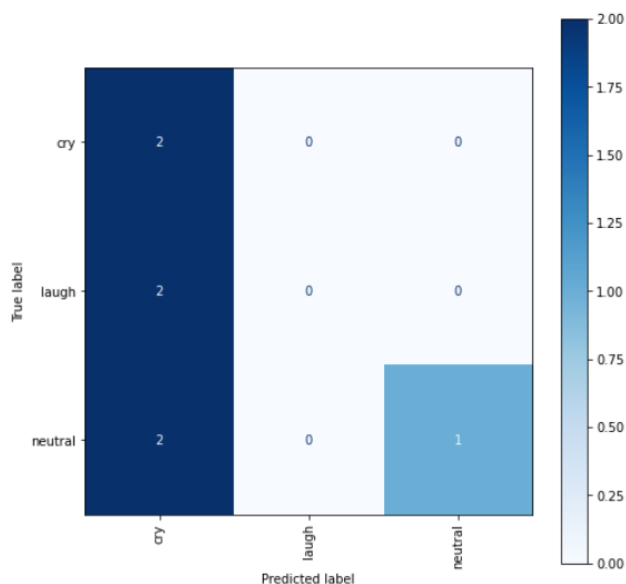
Prosečna tačnost po klasi sada iznosi 74%, a osetljivost je znatno manja, 58%, što je direktno uzrokovano izrazito niskom osetljivošću klase *laugh*, 13%, čije uzorke posledično klasifikator najčešće pogrešno klasifikuje.

S obzirom da su vrednosti mera uspešnosti po klasi donekle srazmerne brojnosti klasa zastupljenim u skupu podataka, uspešnost klasifikacije neuronske mreže bi mogla dodatno da se poboljša primenom neke od metoda za prevazilaženje problema nebalansiranosti skupova (npr. *upsampling*)

| Algoritam | Udeo ispravno pogodjenih klasa |
|----------------------|--------------------------------|
| Logistička regresija | 57% |
| KNN | 43% |
| Neuronske mreže | 29% |

V ZAKLJUCAK

Mali broj uzoraka se javlja jos u pocetnoj bazi podataka, sto nam govori o mogucnosti da model nije imao dovoljno vremena da se obuci i upozna veci broj specifičnosti podataka. Pri podeli na trening i test skup, ne moze se uticati na to u kojem skupu ce se naci specifični uzorci, te pri obucavanju trening skupa, moguće da neki specifični uzorci upadnu i ostanu u test skupu, pa uprkos dobrim merama uspesnosti, ne predviđaju tačne uzorke, ili obrnuto - predviđa dobro a lose su mere uspesnosti, mozda su onda specifični uzorci ostali u trening skupu.



Slika 4. Rezultati neuronske mreže na test skupu

IV REDUKCIJA DIMENZIONALNOSTI

Redukcija dimenzionalnosti je ništa više nego redukcija obeležja. Svrha redukcije dimenzionalnosti jeste da smanji bazu kako bi lakše mogla obučiti model a pritom da se preformanse naglo ne pogoršaju. S, druge strane, treba voditi računa o tome kako se vrši redukcija, da ne bi došlo do prevelikog gubitka informacija.

PCA je jedan od dva ključna algoritma kojim smo vrsili redukciju dimenzionalnosti. PCA algoritam uz prethodnu standardizaciju podataka s ciljem da se zadrže komponente koje obuhvataju 90% ukupne varijanse, početni prostor su od 75 obeležja redukovali na 8. Rezultati pokazuju da obukom modela kroz logisticku regresiju je postignut najveći udeo ispravno klasifikovanih uzoraka, nešto malo uspesnije to cini KNN algoritam, dok neuronske mreže skoro pa potpuno mase tačno predviđanje uzoraka.

