

Klasifikacija instrumenata na osnovu audio zapisa

Vladimir Gajčin, SW-49/2014
Fakultet tehničkih nauka, Novi Sad



Motivacija

Čuli ste Šostakovičev “Valc br. 2”, i zanima vas koje je instrumente koristio?

Ideja ove aplikacije je da nam korisnik prosledi audio zapise kratke dužine (do 20 sekundi) u kojima samo jedan instrument svira, i ostavi algoritmu da odradi ostalo - da prepozna koji je to instrument.

Podaci

- IRMAS skup podataka (sadrži 9579 audio zapisa - 6705 za treniranje, 2874 za testiranje)
- Ručno sastavljen skup za demonstraciju

Podaci za treniranje su raspoređeni i unapred označeni jednim instrumentom.

Podaci za testiranje nisu označeni, ali su im priključeni tekstualni fajlovi sa oznakama instrumenata koje sadrže. U podacima za testiranje postoje audio zapisi koji sadrže i dva instrumenta.

Instrumenti koje ovaj skup podataka sadrži i razlikuje su čelo, klarinet, flauta, akustična gitara, električna gitara, orgulje, klavir, saksofon, truba, violina, i ljudski glas (pevanje). Ima ih 11.

Pretprocesiranje

Kako bi znali po čemu da razlikujemo audio zapise, potrebno je naći njihova karakteristična obeležja.

Za to, korišćeni su *MFCCs (Mel-Frequency Cepstral Coefficients)*. MFC koeficijenti opisuju oblik krivulje spektralne omotnice.

Prvih 20 koeficijenata je odabrano kao obeležje audio zapisa.

Klasifikacija

KNN (K-Nearest Neighbours) algoritam je korišćen za klasifikaciju instrumenata.

Ulazne podatke za treniranje predstavljali su *MFC* koeficijenti audio datoteka iz trening skupa podataka i klase (instrumenti) dobijene iz označenih datoteka.

Ulazne podatke za testiranje predstavljali su *MFC* koeficijenti audio datoteka iz testnog skupa podataka.

Na osnovu ulaznih podataka, radi se predikcija klase (instrumenta), i to predstavlja izlazne podatke testiranja.

Parametar *K* je postavljen na 97, što znači da ćemo za svaki ulazni podatak testiranja posmatrati njegovih najbližih 97 komšija. Klasa koja je najviše prisutna unutar tih 97 komšija predstavlja predikciju klase za svaki ulazni podatak.

Rezultati

Za evaluaciju rešenja korišćeni su anotirani tekstualni fajlovi prikazani uz svaki podatak iz testnog skupa.

Tekstualni fajlovi sadrže oznaku jednog ili dva instrumenta prisutnih u istoimenom audio zapisu.

Rezultati predikcije nad testnim skupom su poređeni sa podacima dobijenim iz tekstualnih fajlova. Tačan rezultat znači da je rezultat predikcije jednak jednom (ili jednom od ukupno dva) instrumentu iz anotiranog tekstualnog fajla.

Za različite *K* parametre, rešenje dostiže određenu tačnost:

K	Tačnost
5	31%
11	34%
31	38%
97	40%

Zaključak

Rešenje raste povećavanjem parametra *K*. Međutim, daljim povećavanjem istog parametra ne bi se primetila velika razlika u tačnosti.

Radi postizanja veće tačnosti, alternative su drugi algoritam za klasifikaciju, povećanje broja *i*/ili kvaliteta obeležja, kao i unakrsna validacija.