

Prepoznavanje zvokov cajona

Vmesno poročilo

1 Problem

Rad bi naredil sistem, ki posluša igranje cajona¹, v realnem času prepozna zvoke udarcev in glede na zaznave predvaja nove sintetične zvoke. Ta problem sem razdelil na dva dela:

1. zaznava udarcev
2. klasifikacija zvokov

Prvi del sem že rešil z uporabo znanega algoritma za zaznavo udarcev (*onset beat detection*) [2, 3]. Drugemu delu pa se bom posvetil v tem projektu. Cajon lahko ustvari dva različna zvoka. Če igramo po spodnjem delu inštrumenta dobimo nižji zvok (kick), če pa igramo po zgornjem delu pa dobimo višji zvok (snare). Poleg tega se zvoki razlikujejo glede na moč udarca.

Cilj tega projekta je razviti model, ki bo zvoke klasificiral v dva razreda (kick in snare). Ta problem postane težji, ko dodamo zahtevo, da mora model delovati v (skoraj) realnem času. To pomeni, da lahko za klasifikacijo uporabi le majhen del zvočnega signala (1 do 2 ms). Tako nizka zakasnitev je potrebna, da lahko med igranjem dodajamo druge zvoke ne da bi igralec opazil zakasnitev.

2 Pristop

2.1 Orodja

Odločil sem se za uporabo programskega jezika Python, saj ima veliko zbirko knjižnic za obdelavo podatkov in strojno učenje. Za ta projekt sem uporabil oziroma nameravam uporabiti (med drugim) naslednje knjižnice:

- **scipy** - za branje in obdelavo zvočnih signalov
- **numpy** - za delo z večdimenzionalnimi seznamami

¹**cajon** [kahón] - tolkalo v obliki lesene resonančne škatle, na kateri se sedi in udarja po sprednji stranici, navadno z roko, lahko z različnimi udarjalkami, v notranjosti ima lahko napeto žico, mrežico, z dodanimi kraguljčki, značilno za perujsko glasbo, v zadnjem času tudi za špansko glasbo, uporablja se ga tudi kot priročno alternativo klasičnim bobnarskim kompletom

- **scikit-learn** - za klasične metode strojnega učenja
- **tensorflow** - za delo z nevronskimi mrežami
- **matplotlib** - vizualizacija signalov in rezultatov

2.2 Učna množica

Za testiranje različnih klasifikacijskih modelov je potrebna čim večja podatkovna množica. Za začetek sem ustvaril podatkovno množico iz zvokov svojega cajona. Za vsak razred (*kick*, *snare*) in štiri različne jakosti (*piano*, *mezzo piano*, *mezzo forte*, *forte*) sem posnel 60 zvokov (vse skupaj 480 zvokov) in jih s pomočjo algoritma za zaznavo udarcev razdelil na posamezne vzorce. Ker je ta množica relativno majhna, sem pri nekaterih pristopih uporabil pristop bogatenja podatkov (data augmentation) in z dodajanjem šuma, spremembo jakosti, frekvenca, hitrosti in zamika ustvaril nove vzorce.

2.3 Določanje značilk

Rad bi preveril različne značilke:

- **statistične značilke** kot so min, max, varianca, povprečna vrednost, momenti, ... na surovem signalu
- **spektralne značilke**, ki izvirajo iz frekvenčnega spektra signala pridobljenega s pomočjo hitre Fourierove transformacije [1].
- **surov signal** (brez značilk) modeli globokega učenja bodo morda lahko sami izluščili značilke iz surovega signala

2.4 Modeli za klasifikacijo

Ker bi rad, da model deluje v realnem času (in morda celo na mikromilniku), bi rad poiskal čim bolj preprost model, ki bo dovolj natančen. Pričakujem, da bo najboljši model predstavljal kompromis med natančnostjo in časovno zahtevnostjo. Nameravam preizkusiti naslednje modele:

- **PCA**: predvsem za vizualizacijo podatkov in preverjanje, če kaka cenilka že sama po sebi dobro loči razrede
- **logistična regresija**: preprost model za klasifikacijo dveh razredov
- **k-najbližjih sosedov**: napoveduje razred glede na najbližje točke v prostoru značilk

- **naključni gozd**: ansambelska metoda, ki združuje več odločitvenih dreves
- **gosta nevronska mreža**: večji model, ki se bo morda bolje prilagodil podatkom
- **konvolucijska nevronska mreža**: model, ki lahko sam izlušči značilke iz surovega signala

3 Nadaljnji koraki

Zaenkrat sem preizkusil nekaj osnovnih modelov in sicer logistično regresijo, PCA, in gosto nevronske mreže. V naslednjem koraku bom sistematično preizkusil različne kombinacije značilk in klasifikacijskih modelov. Izmeril bom njihovo natančnost na (prej ne videni) testni množici in ocenil časovno zahtevnost.

Literatura

- [1] William Brent. Cepstral analysis tools for percussive timbre identification. 01 2009.
- [2] P Brossier, JP Bello, and MD Plumbley. Real-time temporal segmentation of note objects in music signals, 2004.
- [3] Fredrik Gustafsson. *Adaptive Filtering and Change Detection*, chapter On-line Approaches. 2001.