

Predviđanje nivoa depresije kod ljudi na osnovu audio zapisa

Analiza akustičkih karakteristika govora za detekciju nivoa depresije

Veselin Roganović, SV 36/2022 | Fakultet tehničkih nauka, Novi Sad



Uvod

Depresija predstavlja jedan od najvećih izazova modernog društva. Pravovremena detekcija je ključna, ali često teška čak i za stručnjake. Ovaj projekat istražuje mogućnost automatske procene nivoa depresije na osnovu **akustičkih karakteristika govora**, bez analize sadržaja izgovorenih reči.

Skup podataka

- Korišćen je **DAIC-WOZ** skup podataka koji sadrži:
- Audio zapise kliničkih intervju.
 - PHQ-8** skorove (0-24) kao ciljnu vrednost.
 - Ukupno 140 uzoraka (80% trening, 10% val, 10% test).

Pretprocesiranje

- Resampling na 16kHz i normalizacija amplitude.
- Ekstrakcija MFCC koeficijenata (u vremenskim prozorima).

Metodologija

Primenjena su dva različita pristupa:

1. SVR (Support Vector Regression):

- Ulaz: MFCC matrice (vremenski prozori).
- Arhitektura: SVR sa RBF kernelom ($C=1$, $\epsilon=0.01$, $\gamma=\text{'scale'}$)

2. CNN + LSTM:

- Ulaz: MFCC matrice (vremenski prozori).
- Arhitektura: Konvolucionni slojevi za prostorna obeležja praćeni LSTM slojevima za sekvence i potpuno povezanim slojem.

Arhitektura CNN+LSTM

- CNN:** 2D slojevi sa BatchNormalization i Dropout-om.
- LSTM:** 2 sloja sa 64 skrivene jedinice.
- Gubitak:** Kombinovani MAE i Pearson loss.

Rezultati

Validation

Model	MAE	RMSE	Prsn
SVR	4.29	4.98	0.52
CNN+LSTM	3.99	5.02	0.45

Test

Model	MAE	RMSE	Prsn
SVR	4.37	6.17	0.27
CNN+LSTM	5.51	7.20	0.05

SVR pobeđuje u svim kategorijama, osim za MAE na validacionom skupu.

Zaključak

Rezultati potvrđuju da akustička obeležja nose značajne informacije o mentalnom stanju govornika.

Korišćeni alati

Python, PyTorch, Librosa, Scikit-Learn