

Real time klasifikácia emocionálneho stavu z reči

Potenciál využitia

Hlavnou motiváciou pokusov o strojové rozpoznávanie emócií sú široké možnosti využitia v prípade úspechu. Medzi oblasti kde by rozpoznávanie emócií mohlo nájsť uplatnenie patrí napríklad medicína, núdzové linky, inteligentné výukové systémy, komunikácia so zákazníkom v call centrách, rozšírenie možností interakcie s počítačom a rôzne formy umelej inteligencie.

Charakteristiky reči

Zo signálu reči je možné vypočítať rôzne charakteristiky (frekvencia, intenzita, mel frekvenčné kepstrálne koeficienty atď). Charakteristiky je možné vnímať ako globálne štatistiky z celého zvukového súboru (napr. minimum, maximum, priemer) alebo lokálne hodnoty (vektor hodnôt charakteristiky v čase). Na výpočet charakteristík boli použité Python knižnice librosa a parselmouth.

Nahrávky emócií

Schéma definovaná psychológom Paulom Ekmanom obsahuje 7 základných emócií – hnev, znechutenie, strach, šťastie, smútok, prekvapenie a neutrálnu emóciu. Nahrávky emócií použité na tréning a testovanie neurónových sietí boli získané z databáz RAVDESS, EMO-DB a SAVEE. Problémom týchto databáz je nízky počet nahrávok, neuniformné triedy a fakt, že boli vytvorené umelo prostredníctvom hercov, avšak lepšia verejne dostupná databáza doposiaľ chýba.

Predspracovanie vstupných súborov

Spracovanie pred extrakciou charakteristík rieši alebo zmiernuje niektoré problémy použitých

databáz. Zahŕňa vyrovnanie uniformity tried, orezanie ticha a normalizáciu. Dôležitým krokom je takisto augmentácia – zväčšenie trénovacej vzorky umelým vytvorením nových súborov úpravou existujúcich.

Experimenty s neurónovými sieťami

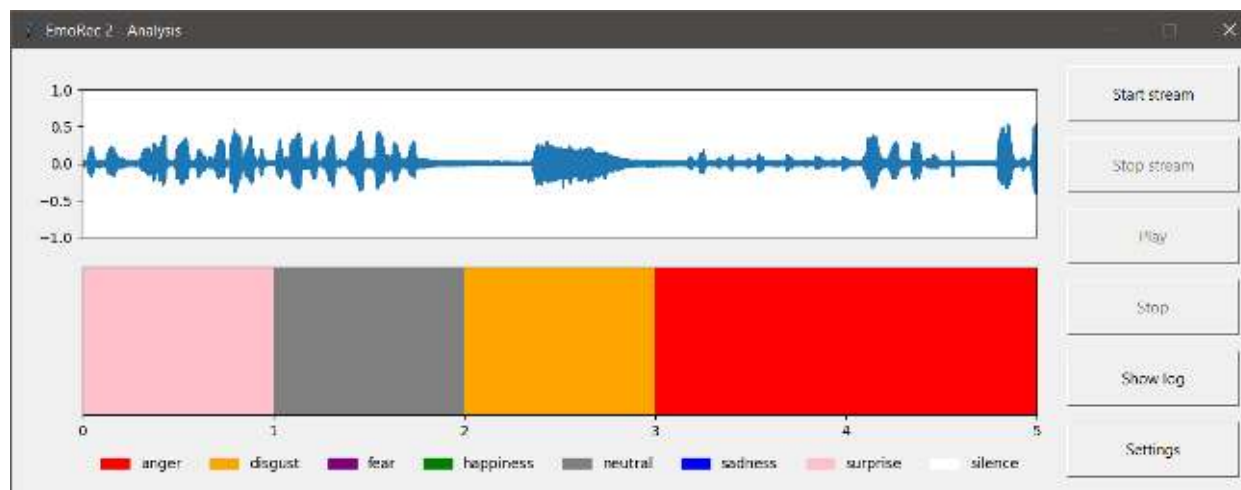
Dôležitou súčasťou výskumu strojového rozpoznávania emócií je hľadanie vhodného typu klasifikátora a takisto vhodnej kombinácie charakteristík, ktoré sa stanú jeho vstupom. V práci boli realizované 3 experimenty:

- 1) Dopredná neurónová sieť so 45 globálnymi charakteristikami na vstupe
- 2) 1D konvolučná sieť s vektormi lokálnych hodnôt energie, zerocrossingu a 13 mfcc
- 3) 2D konvolučná sieť, ktorej vstupom sú spektrogramy

Pri výbere najlepšieho modelu a metodiky bolo navyše nutné brať ohľad aj na čas, za aký je možné extrahovať a spracovať charakteristiky zo vstupného signálu, a ako dlho trvá následná predikcia emócie. Ako najvhodnejší sa ukázal model z experimentu 2, ktorý s testovacím subsetom dosiahol 78,85% úspešnosť klasifikácie a navyše bol veľmi rýchly.

Program EmoRec 2

Výsledkom práce je Python aplikácia s grafickým rozhraním, ktorá dokáže v reálnom čase spracovávať prúd údajov z vybraného vstupného zariadenia (napr. z pripojeného mikrofónu). Vstup je segmentovaný na krátke úseky, z ktorých sú vypočítavané charakteristiky. Model neurónovej siete z experimentu 2 potom vykoná predikciu emócie. Model aj aplikácia poskytujú priestor na vylepšovanie a potenciálne využitie v praxi.



Rozhranie aplikácie EmoRec 2