

Identificarea Emoțiilor din Semnale Audio

Introducere

Cu avansarea tehnologiei și tot mai multe aplicații care interacționează cu utilizatorii prin intermediul sunetului, recunoașterea emoțiilor din fișiere audio devine o componentă crucială în dezvoltarea sistemelor inteligente și a interfețelor utilizator intuitive. Proiectul se concentrează pe utilizarea setului de date TESS Toronto emotional speech pentru a construi un model de recunoaștere a emoțiilor, capabil să identifice și să clasifice emoțiile exprimate în discursul audio.

Descriere a Componentelor Proiectului

Biblioteci Python

- **pandas**: Folosită pentru manipularea și analiza datelor, în special pentru crearea și gestionarea datelor într-un format tabular.
- **numpy**: Bibliotecă de calcul numeric pentru operații matematice eficiente.
- **os**: Furnizează funcții pentru interacțiunea cu sistemul de operare, în acest caz pentru navigarea prin directoarele fișierelor audio.
- **matplotlib.pyplot**: Utilizată pentru a genera grafice și vizualizări, precum graficele de undă și spectrele de frecvență ale semnalului audio.
- **tqdm**: Oferă o bară de progres pentru a monitoriza progresul în timpul procesului de calculare a coeficientilor MFCC .
- **librosa.display**: Folosită pentru afișarea spectrogramelor și altor reprezentări ale semnalului audio.
- **librosa.feature**: Folosită pentru extragerea caracteristicilor semnificative ale semnalului audio, cum ar fi coeficienții MFCC.
- **tensorflow.keras.models.Sequential**: Interfața secvențială pentru definirea și antrenarea modelelor de rețele neuronale în TensorFlow.
- **tensorflow.keras.utils.to_categorical**: Folosită pentru a realiza codificarea one-hot a etichetelor claselor în timpul antrenării rețelei neuronale.

Setul de Date TESS Toronto Emotional Speech:

- **Descriere:** Setul de date conține fișiere audio cu discurs uman ale unui barbat și unei femei, exprimând diferite emoții precum bucurie, tristețe, frică etc. Fiecare fișier este etichetat corespunzător cu emoția pe care o exprimă. Fiecare emoție are un total de 400 de fișiere audio asociate
- **Utilizare:** Setul de date este utilizat pentru antrenarea și evaluarea modelului de recunoaștere a emoțiilor. Fiecare fișier audio este încărcat, iar caracteristicile sale, în special coeficienții MFCC, sunt folosite pentru a antrena rețeaua neurală pentru a clasifica emoțiile.

Funcții:

- **load_data():** Navighează prin directoarele fișierelor audio și creează un DataFrame Pandas pentru a gestiona căile fișierelor și etichetele asociate acestora.
- **extract_mfccs(dataframe):** Încarcă datele audio, calculează coeficienții MFCC și le returnează sub forma unui vector de matrice.
- **extract_emotions(data, sampling_rate, model, emotion="unknown"):** Extrage caracteristici dintr-un fișier audio, îl prezintă grafic, și utilizează un model pentru a prezice și afișa emoția asociată.
- **preprocess_mfccs(mfccs, max_length):** Realizează preprocesarea secvențelor de coeficienți MFCC pentru a le aduce la o lungime fixă, necesară pentru antrenarea rețelei.

Cum putem să clasificăm emoțiile:

Din mai multe studii, dar și în funcție de rezultatele din setul nostru de date, observăm caracteristicile asociate emoției tristețe au un "feature distance" mai mare între ele comparativ cu celelalte emoții. Următorul "feature distance" mai mare este asociată cu starea emoțională neutră. Starea emoțională de furie prezintă o distanță relativ mai mică, iar emoția de bucurie are cea mai mică distanță între caracteristici. Se folosește termenul de "feature distance" pentru a descrie diferența sau distanța dintre caracteristicile (cuvintele rostite) asociate diferitelor emoții. Aceste caracteristici ar putea fi atribuțiile extrase din semnalul audio, cum ar fi coeficienții MFCC.

Coeficienții MFCC (Mel-Frequency Cepstral Coefficients):

Procesul de extragere a acestor coeficienți se desfășoară în mai multe etape:

1. **Filtrarea Mel (Mel-Frequency Filterbank):** Semnalul audio este divizat în ferestre mici de timp, iar pentru fiecare fereastră se aplică un filtru Mel, care este proiectat pentru a imita

sensibilitatea urechii umane la diferite frecvențe. Acest proces produce o serie de benzi de frecvență Mel.

2. **Transformata Cepstrală:** Se aplică transformata cepstrală pe spectrul de putere al semnalului, obținut prin calculul spectrului de putere al semnalului filtrat Mel. Acest pas ajută la evidențierea caracteristicilor semnalului în domeniul cepstral.
3. **Selectarea Coeficienților Cepstrali:** Din rezultatul transformatei cepstrale, se selectează un număr specific de coeficienți (de obicei, primele 13-20) pentru a reprezenta semnalul. Primii coeficienți dețin cele mai multe caracteristici din semnalul audio, de aceea nu este relevant câți coeficienți se aleg, totuși în acest proiect au fost folosiți primii 20.

Deci, coeficienții MFCC reprezintă o formă compactă a caracteristicilor semnalului audio, capturând aspecte importante ale spectrului de frecvențe și ale modulațiilor temporale ale semnalului vocal.

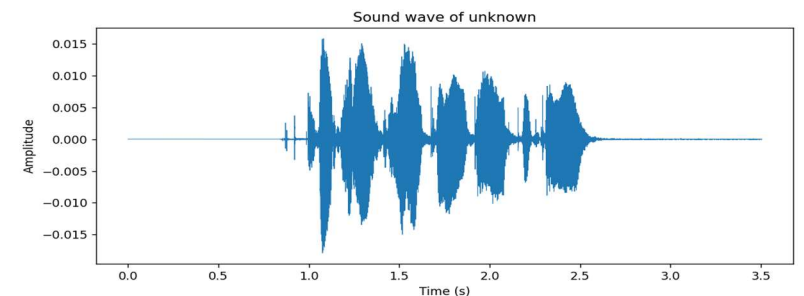
Diferențele și tendințele între coeficienții MFCC pot furniza informații despre variațiile semnificative ale caracteristicilor semnalului în timp, ajutând la capturarea modificărilor de ton, ritm și alte aspecte ale vorbirii.

Rezultate ale clasificării unui semnal audio după emoții:

În acest proces, am plotat următoarele grafice:

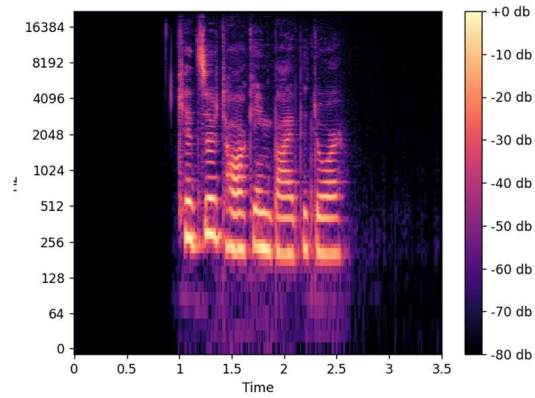
SoundWave Plot:

- Plotarea undei sonore în funcție de timp. Acest grafic oferă o reprezentare a modului în care amplitudinea semnalului variază pe parcursul timpului. Util pentru a vizualiza forma generală a semnalului audio.



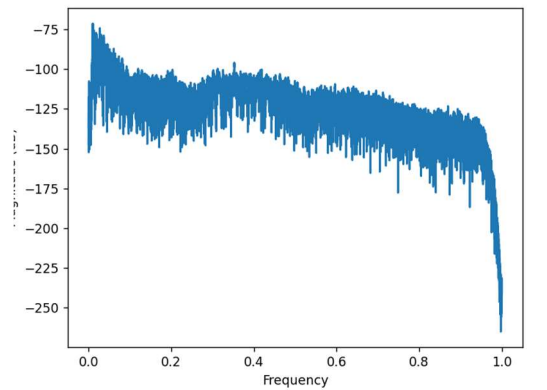
Logarithmic Spectrogram Plot:

- Plotarea spectrogramului într-o scală logaritmică. Acest lucru evidențiază mai bine detaliile în frecvențe mai joase și face mai vizibile evenimentele subtile în spectru.



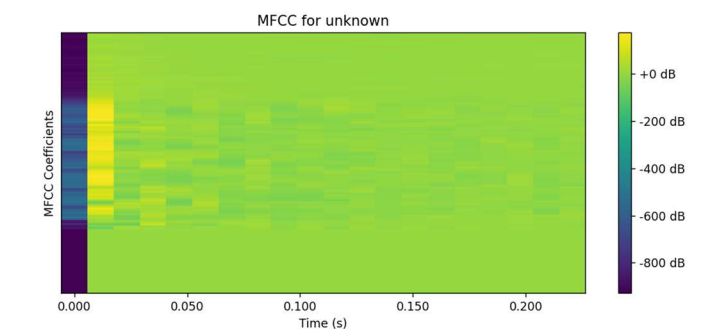
Magnitude Spectrum Plot:

- Plotarea spectrului de magnitudine. Acest grafic arată intensitatea semnalului în funcție de frecvență.



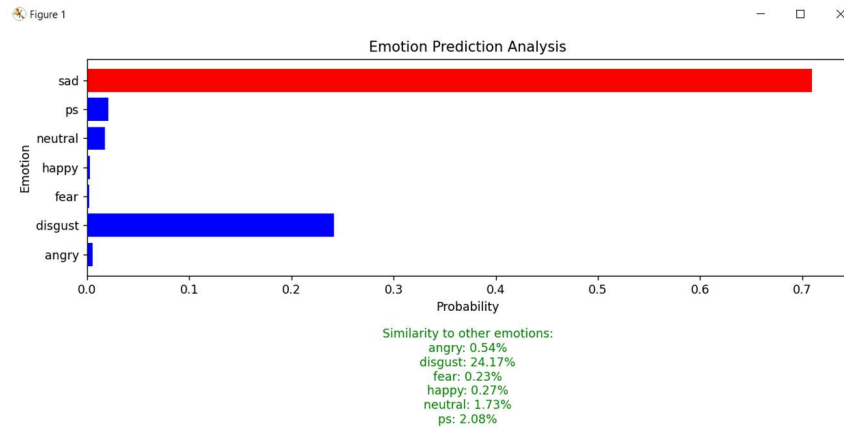
MFCC Coefficients Plot:

- Plotarea coeficienților MFCC (Mel-Frequency Cepstral Coefficients)



Prediction Bar Chart:

- Plotarea unui diagramă de bare care arată probabilitatea prezisă pentru fiecare clasă de emoții.



Concluzie:

Acest proiect evidențiază procesul de recunoaștere a emoțiilor din datele audio folosind tehnici de învățare automată. Funcțiile implementate facilitează încărcarea datelor, extragerea de atribute, crearea modelului și extragerea emoțiilor, oferind o soluție cuprinzătoare pentru analiza emoțiilor în fișierele audio. Insa, calitatea predicțiilor poate fi influențată de calitatea datelor de antrenare.