Relazione per tirocinio ***Falconi Eleonora***

Feature utilizzate

- MFCC: è la caratteristica spettrale più utilizzata nel riconoscimento delle emozioni del parlato. È una caratteristica vantaggiosa per raccogliere i benefici dell'audio grezzo. Funziona molto efficacemente anche in presenza di rumore. In questo studio sono state incorporate 39 caratteristiche di MFCC.
- Mel-spectrogram: lo spettrogramma in ingresso viene mappato direttamente sulla funzione base Mel. La rappresentazione dello spettrogramma di Mel aumenta la chiarezza uditiva del sistema.
- Chroma: le caratteristiche di Chroma sono strettamente collegate alle 12 diverse classi di altezza. Il chroma viene utilizzato per catturare le caratteristiche armoniche e melodiche della musica e del suono.
- Contrast: è l'energia media stimata di un suono. Quindi un basso valore di contrasto corrisponde a un rumore a banda larga.
- Tonnetz: questo tipo di caratteristica funziona in modo simile al chroma.

Pre-processazione dei dati

Per aumentare il database relazionale sono state applicate diverse tecniche di aumento dei dati, come ad esempio il pitch shifting e il time stretching. Quest'ultimo consiste nel modificare la velocità del suono con alcuni parametri, mentre il pitch shifting è una variazione di frequenza della nota musicale, ridotta o aumentata di una certa quantità.

Il modello di algoritmo proposto è basato sul modello della rete neurale a convoluzione(CNN). I modelli CNN sono in grado di estrarre automaticamente caratteristiche dall'input; quindi, questo modello di convoluzione unidimensionale prende direttamente tutti i file audio e li memorizza in un array per poi mettere a punto le caratteristiche sperimentali utili. Mentre per quanto riguarda il modello di apprendimento profondo multilingue K-fold(MDLM) i dati standard e quelli aumentati che vengono utilizzati, vengono suddivisi in set di addestramento e set di test. Quindi per prima cosa vengono presi i dati di addestramento, i quali vengono inseriti in un modello con convalida incrociata k-fold e se il modello funziona anche con i dati espansi, l'algoritmo verrà addestrato con i dati di test originali. Quindi l'obbiettivo finale è quello di rilevare emozioni multilingue dal file di tipo grezzo. La validazione incrociata a k-fold è una tecnica utilizzata nell'apprendimento automatico per valutare le prestazioni di un modello su un insieme di dati limitato. Poiché generalmente l'insieme di dati originale viene diviso in k parti di uguali dimensioni(o anche "fold"). Il modello verrà addestrato su k-1 fold dei dati e poi testato sull'unico fold rimanente e questo processo viene ripetuto k volte utilizzando ogni volta una diversa parte come set di test e le rimanenti come set di addestramento.