



ESPEECHPY

EMOTION SPEECH RECOGNITION

PROGETTO REALIZZATO

DA:

CLAUDIO PACE

GIANFRANCO SEMERARO

MATTEO MARIANO

MICHELE METTA

NICOLA NARGISO

PIETRO SANGERMANO


DOCENTE DEL
CORSO

Prof.ssa DE CAROLIS



COS'È L'EMOTION SPEECH RECOGNITION?

E' il riconoscimento delle emozioni
attraverso l'analisi della voce di un
individuo in base a delle caratteristiche
(features) prestabilite



Cos'è un agente?

E' un modulo software che rispetta i seguenti attributi:

- **Autonomia:** capacità di controllare i propri comportamenti senza comportamenti esterni;
- **Abilità Sociale:** capacità di comunicare con altri agenti umani e non;
- **Reattività:** capacità di fornire risposte immediate ai cambiamenti ambientali che vengono percepiti;
- **Proattività:** capacità di prendere iniziativa.

Tutto questo grazie ad **organi sensori**, **organi attuatori** e **razionalità** per raggiungere l'obiettivo prestabilito.

Cos'è un classificatore?

Un classificatore è un algoritmo basato su un modello matematico in grado di apprendere, dalle caratteristiche di un oggetto, la classe (o gruppo) di appartenenza.

Le caratteristiche di un oggetto sono anche conosciute come *valori di feature* e sono tipicamente presentate alla macchina in un vettore chiamato *vettore di feature*.

Classificatori a confronto

Il confronto, fra classificatori, viene effettuato in base all'accuratezza indicata dal "test_score".

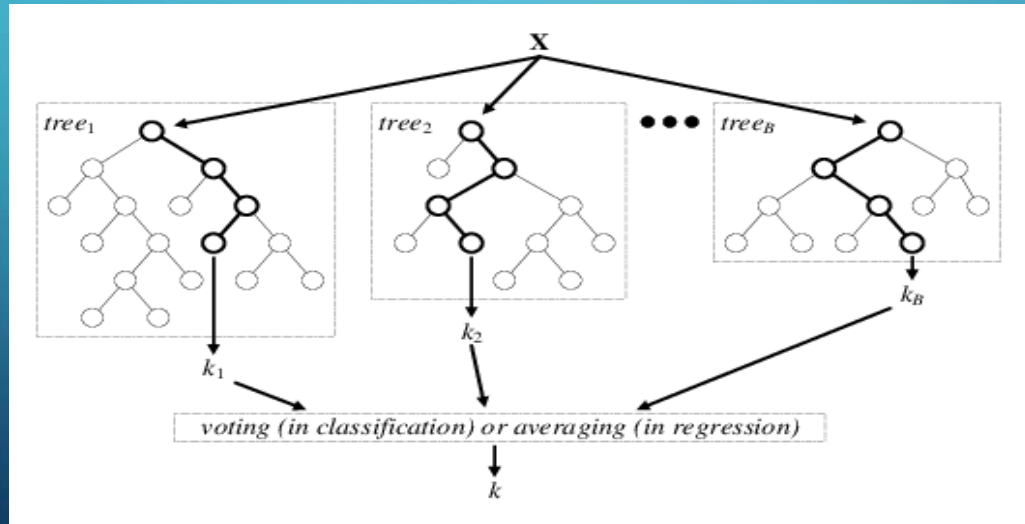
Nel nostro caso, il Random Forest ha il punteggio più alto.

	classifier	train_score	test_score	train_time
5	Random Forest	1.000000	0.672316	1.989251
3	Gradient Boosting Classifier	1.000000	0.627119	5.056969
4	Decision Tree	1.000000	0.531073	0.004600
0	Logistic Regression	0.401460	0.389831	0.018028
6	Random Forest Regressor	0.907852	0.347215	3.225850
8	Naive Bayes	0.391727	0.327684	0.001030
1	Nearest Neighbors	0.520681	0.305085	0.000628
7	Neural Net	0.236010	0.276836	0.193631
2	Linear SVM	1.000000	0.107345	0.024546

	classifier	train_score	test_score	train_time
5	Random Forest	1.000000	0.539792	2.850700
3	Gradient Boosting Classifier	1.000000	0.460208	9.308987
4	Decision Tree	1.000000	0.380623	0.011816
0	Logistic Regression	0.323442	0.318339	0.024586
8	Naive Bayes	0.298220	0.314879	0.001274
2	Linear SVM	1.000000	0.252595	0.061609
6	Random Forest Regressor	0.886688	0.211917	6.016947
1	Nearest Neighbors	0.405045	0.211073	0.000874
7	Neural Net	0.152819	0.155709	0.145770

Classificatore Random Forest

Il Random Forest è un classificatore d'insieme che è composto da alberi di decisione. Fornisce come risultato la classe che corrisponde agli output degli alberi; I vari risultati vengono prese in considerazione in base ad un criterio di scelta, come il majority voting



Due modelli a confronto

Testando diversi modelli di classificatori (da noi serializzati) , siamo riusciti a creare 2 modelli che risultano essere migliori:

- ***Modello3.pickle***
- ***Modello5.pickle***

Questi sono stati addestrati esclusivamente su dataset “EMOVO” perché, integrando anche il nostro dataset, avremmo raggiunto una minore percentuale di accuratezza: nella fase di test, abbiamo rilevato un riscontro negativo nelle predizioni eseguite dal sistema.

Metriche di Validazione

Per valutare i due modelli, abbiamo utilizzato le metriche:

- **MSE**(MeanSquaredError)
 - L'**Errore Quadratico Medio** indica la media dei quadrati degli errori fra valori stimati e valori reali
- **RMSE**(RootMeanSquaredError)
 - La **Radice dell'Errore Quadratico Medio** corrisponde alla varianza interna data dal rapporto fra devianza interna e numerosità totale
- **MAE**(MeanAbsoluteError)
 - La **Media degli errori assoluti** indica la media fra le differenze assolute tra previsione e valore vero
- **Matrice di Confusione**
 - restituisce una rappresentazione dell'accuratezza di

MAE & RMSE & MSE

Sono metriche per confrontare le previsioni con i risultati finali

Modello3.pickle

```
Mean Absolute Error: 1.08843537414966  
Mean Squared Error: 3.707482993197279  
Root Mean Squared Error: 1.9254825351576885
```

Modello5.pickle

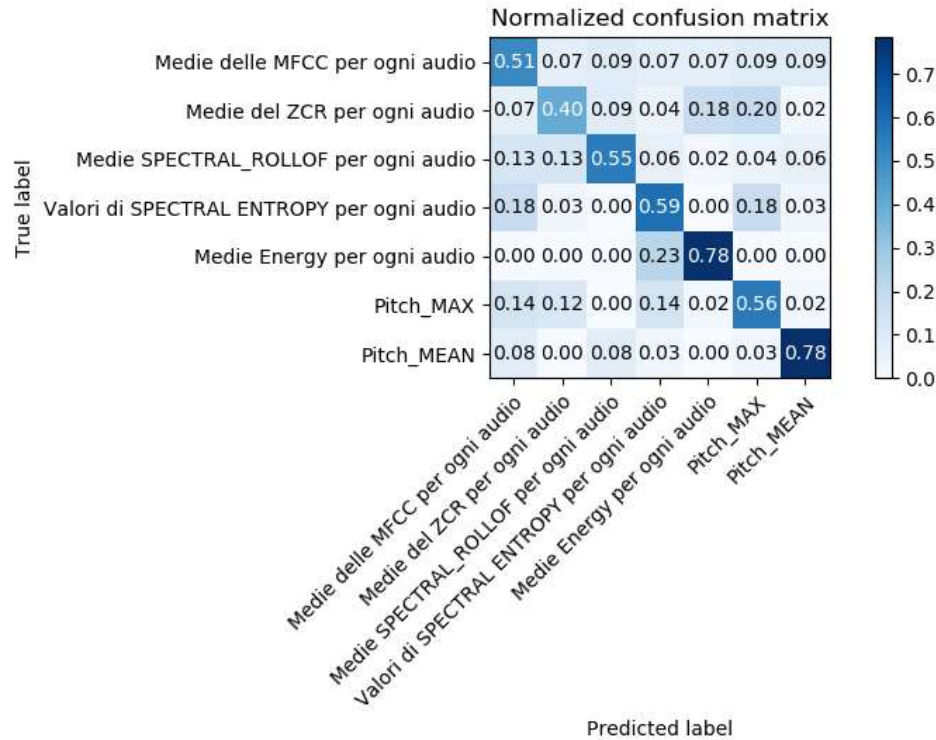
```
Mean Absolute Error: 1.5746887966804979  
Mean Squared Error: 5.2012448132780085  
Root Mean Squared Error: 2.2806237772324502
```

Matrice di Confusione

Le colonne della matrice rappresentano i valori predetti. Gli elementi (x,y) sulla diagonale rappresentano la percentuale di casi in cui il classificatore ha catalogato le features in maniera corretta. Attraverso questa matrice, è possibile osservare se vi è "confusione" nella classificazione delle diverse features.

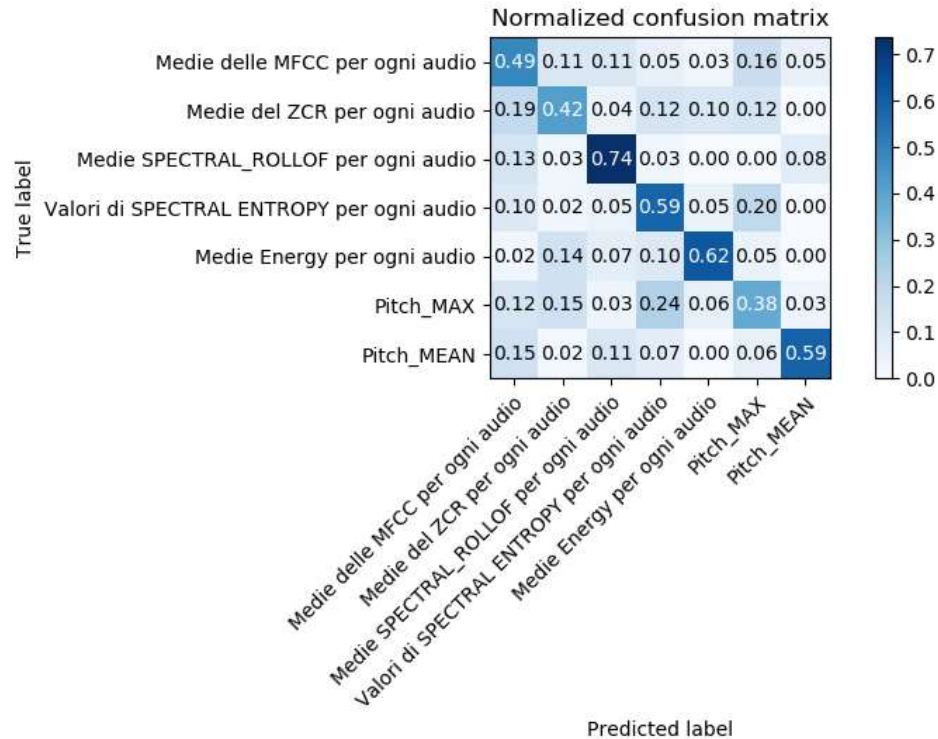
MODELLO3.PICKLE

matrice di confusione



MODELLO5.PICKLE

matrice di confusione



Features

Le *features* utilizzate sono: media degli MFCC, media degli ZCR, media dei Rollof, entropia spettrale, energia, pitch massimo, pitch medio, Intensità minima, media e massima, differenza tra pitch massimo e pitch medio, media derivate degli MFCC.

«FEATURE» PT.1

- **Media degli MFCC:** è la media di tutti i valori MFCC(Mel Frequency Cepstral Coefficients). L'insieme di questi coefficienti costituisce un MFC(**mel-frequency cepstrum**): una rappresentazione a breve termine dello spettro di potenza di un suono;
- **Media degli ZCR:** è la media di tutti i valori ZCR(Zero Crossing Rate). Gli ZCR sono le frequenze del cambiamento del segno lungo il

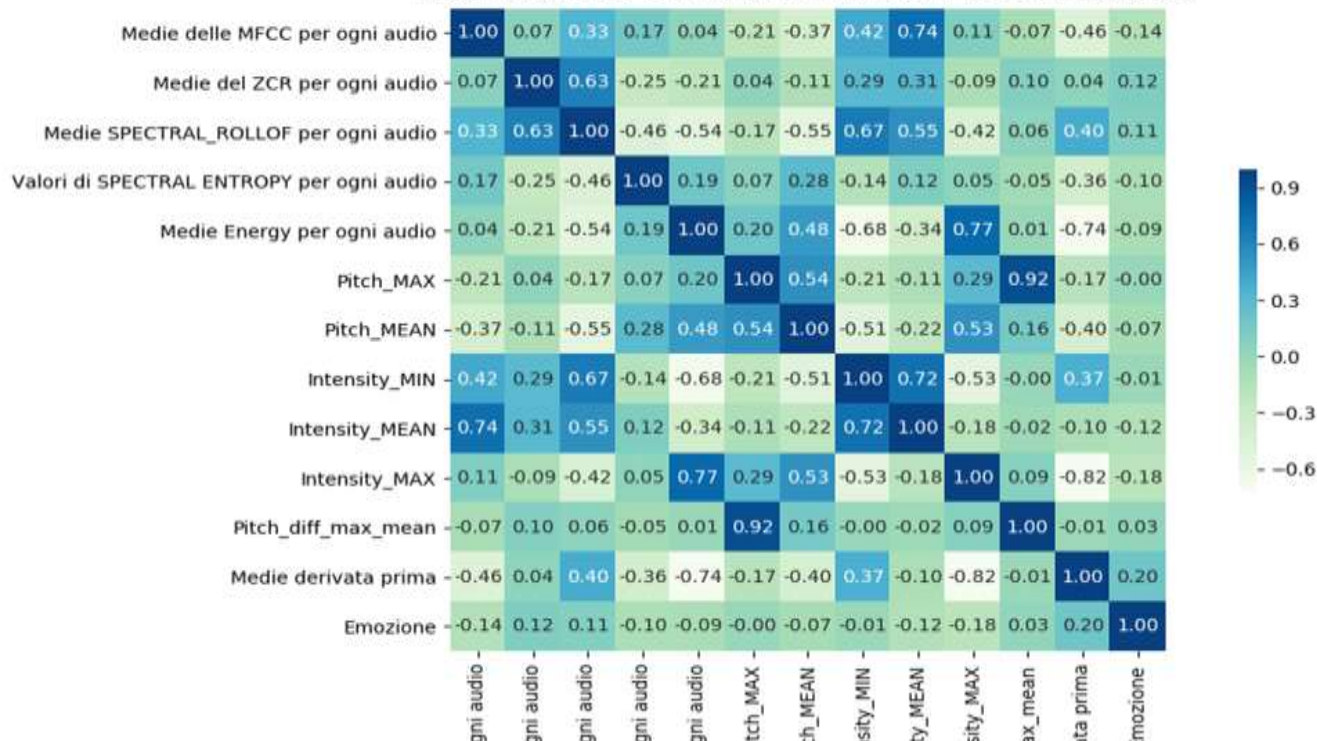
«FEATURE» PT.2

- **Media dei Rollof:** è la media dei valori di Rollof. Questi indicano la porzione di spettro contenente una percentuale alfa (di default 85%) dello spettro di potenza;
- **Entropia Spettrale (o Entropia di Shannon):** definisce la densità dello spettro della potenza (Power Spectral Density (PSD)) dei dati, compone l'arousal e la valence;
- **Energia:** energia meccanica (o cinematica) che, partendo dalla sorgente, viene irradiata sotto forma di onde.

«FEATURE» PT.3

- **Pitch:** l'altezza relativa (o la bassezza relativa) di un tono percepito dall'orecchio umano che dipende dal numero di vibrazioni al secondo prodotte dalle corde vocali;
- **Intensità:** è la qualità acustica e psicoacustica associata alla forza di un suono, determinata dalla pressione esercitata dall'onda sonora.

Correlation matrix between the features



Le emozioni rilevabili attraverso la classificazione

- Rabbia
- Tristezza
- Sorpresa
- Paura
- Disgusto
- Neutralità
- Gioia



Struttura del dataset

Il dataset è composto da un insieme di audio (in formato wav, 48 KHz, 16 bit) suddivisi in base all'emozione che esprimono.



TECNOLOGIE UTILIZZATE



- **Pycharm** - ide della JetBrain
- **Interprete Python 3.7**
- **Librerie utilizzate** : time, asyncio, spade, cv2, numpy, pygame, speech_recognition, pickle, sklearn, pyAudioAnalysis.audioFeatureExtraction, Praat_feature.prosodic_feature, os, librosa, eyed3, matplotlib.pyplot, scipy, pitch, python_speech_features

Interfaccia

L'**interfaccia grafica**, realizzata con l'aiuto di Photoshop e l'Editor Video di Windows 10 comunica con l'utente tramite **DragoBot**, un simpatico draghetto che **reagisce** alle emozioni.



Caratteristiche del nostro agente

Le proprietà principali di “ESpeechPy” sono:

- **Reattività:** risposta ai cambiamenti nell'ambiente
- **Socialità:** capacità di comunicare con umani
- **Continuità Temporale:** costantemente in esecuzione
- **Personalità:** ha una personalità e uno stato emotivo
- **Autonomia:** controllo delle proprie azioni

Funzionamento

Il nostro agente “**ESpeechPy**” rileva le emozioni di un individuo attraverso la voce. In particolare, acquisisce una traccia audio dal microfono del dispositivo che si sta utilizzando e, attraverso le features dell’audio, inviate al classificatore di tipo Random Forest, classifica l’emozione riportando un’animazione video.

Fasi di esecuzione

- Acquisizione della traccia audio da microfono
- Estrazione delle feature dalla traccia audio
- Passaggio delle feature al Random Forest
- Restituzione della classe di appartenenza dell'emozione dell'audio (animazione)

The background is a blue gradient. In the corners, there are decorative white lines resembling circuit traces or a stylized tree structure, with small circles at the end of the lines.

Grazie per
l'attenzione