# Combinazione di tecniche biometriche

**Autore: Saladino Alessio**

# Sommario

# 1. Introduzione

Lo scopo di questo caso di studio è la realizzazione di un sistema biometrico che consenta di eseguire l'autenticazione di un utente usando il volto e la voce.
Per l'addestramento e il test di questo sistema è stato utilizzato un subset del dataset VoxCeleb, che contiene al proprio interno immagini del volto e registrazioni della voce di più di 1000 personaggi famosi.
A causa dei limiti tecnici della macchina su cui è stato fatto girare il sistema, si è scelto di utilizzare solo 15 degli utenti presenti nel dataset.

# 2. Informazioni generali sul software

Il software è stato realizzato in linguaggio Python ed è stato suddiviso in diversi moduli.
I due sistemi realizzati per il riconoscimento del volto e della voce possono essere avviati e testati in maniera indipendente.La combinazione sarà invece svolta da un altro modulo apposito
Sia la parte di riconoscimento del volto che quella di riconoscimento della voce hanno diversi moduli associati:

1. Preprocessing e Feature Extraction
2. Training
3. Tuning
4. Testing

Infine è presente un ultimo modulo chiamato Combination che si occuperà di eseguire la combinazione dei due sistemi.

# 3. Dettagli implementativi

In questa sezione si andranno ad illustrare nel dettaglio le implementazioni delle varie componenti presenti nel sistema.

## 3.1. Preprocessing e Feature Extraction

La prima operazione svolta sulle immagini caricate all'interno del programma consiste nell'eseguire un crop della zona del volto, per farlo si è utilizzata la libreria Python OpenCV.
Successivamente l'immagine con il volto ottenuto viene ridimensionata ad una grandezza di 200x200 e suddivisa in 4 quadranti, per ogni quadrante vengono calcolati gli istogrammi del local binary pattern, gli istogrammi calcolati in ciascun quadrante vengono inseriti in un array che rappresenterà le feature del volto.
Il raggio del LBP utilizzato è 1 e i vicini presi in considerazione sono 8.
Per quanto riguarda la parte di analisi della voce invece sono state scelte come features i Mel Frequency Cepstral Coefficents e le rispettive derivate ottenute da ciascun file wav.
A causa della differenza di lunghezza dei file audio importati, ogni array di features presentava una lunghezza differente, per cui è stata individuata la lunghezza massima e minima di tali array ed è stata eseguita un'operazione di zero padding su tutti gli altri array che non raggiungevano la dimensione massima.
Tutti gli array sono stati suddivisi in frame pari alla lunghezza minima,ciascun frame non nullo è stato trattato come un array di feature.
Ad ogni array di features è stata associata un'etichetta, corrispondente al nome della persona a cui corrispondono le tali features.
Da notare come le cartelle contenenti i file audio all'interno del dataset non siano state nominate con il nome dell'utente(come invece accade per le cartelle contenenti le immagini del volto) ma

con un id (ad esempio l'utente 1 avrà un id pari a id10001),il dataset tuttavia fornisce un file chiamato vox1_meta.csv che viene utilizzato per tradurre l'id dell'utente nel nome o viceversa.
In questo caso è stato scelto di tradurre l'id nel nome esteso dell'utente.
Infine viene eseguito lo split del set di features ottenute in 3 componenti: train set, validation set e test set, con un rapporto di 80% 10% 10%.
Questi set serviranno rispettivamente durante le fasi di training,di tuning e di testing.

## 3.2. Training

Durante questa fase si è preso in considerazione ciascuno degli utenti registrati, e per ognuno degli utenti si sono svolte diverse operazioni:
Le etichette del training set sono state codificate in maniera da valere 1 se l'esempio associato riguarda l'utente per il quale si sta creando il modello, 0 altrimenti.Questo viene fatto allo scopo di dare una classificazione binaria agli esempi che verranno poi mostrati al modello in fase di addestramento.
Il modello scelto è stata una Support Vector Machine avente come parametri un Costo C=10 e un kernel di tipo RBF.
Al termine di questa fase ciascun utente avrà il proprio modello personale che potrà essere utilizzato nelle fasi successive.Il modello sarà in grado di classificare gli esempi che gli verranno mostrati successivamente in due possibili categorie : accettato e rigettato.

## 3.3. Tuning

Questa fase ha lo scopo di individuare un valore di soglia ideale per ogni utente, tale valore viene calcolato utilizzando la curva ROC, ottenuta confrontando i valori di False Positive Rate e di True Positive Rate predette dal modello per il quale si sta cercando di individuare la soglia.Gli esempi utilizzati per il calcolo di tale curva sono quelli salvati all'interno del validation set
Una volta ottenuta una soglia ideale per un utente, essa viene salvata nella stessa cartella del modello all'interno di un file .txt, in modo da poter essere riutilizzata successivamente durante la fase di Testing.

## 3.4. Testing

Quest'ultima fase ha lo scopo di valutare le performance del modello sul test set.La valutazione avviene in maniera separata per ogni utente registrato, del quale viene caricato il modello corrispondente con la rispettiva soglia.Al modello viene fatta predirre la probabilità di ciascun elemento del test set di appartenere ad una delle due classi possibili(accettato o rigettato).Verrà presa in considerazione la probabilità che l'esempio mostrato sia accettato e se tale valore di probabilità supera la soglia associata all'utente attivo allora l'esempio verrà classificato come accettato, altrimenti verrà classificato come rigettato.
Al termine di questa fase si otterranno le matrici di confusione di ciascun utente che saranno in grado di mostrare i Veri Positivi (TP), Veri Negativi (TN), Falsi Positivi (FP) e i Falsi Negativi (FN).Usando questi valori si potrà calcolare l'Accuratezza del modello, che sarà pari a:

**Accuratezza** = (TP+TN)/(TP+TN+FP+FN)

Al termine della fase di testing verrà calcolata l'accuratezza media ottenuta su tutti gli utenti.

# 4. Combinazione

All'interno di questo modulo vengono caricate sia le features riguardanti il volto che quelle riguardanti la voce con le relative etichette.
Per ogni utente verranno svolte diverse operazioni:

- Codifica delle etichette in modo che assumano valore 1 per gli esempi positivi e 0 per gli esempi negativi in base all'utente attivo al momento.
- Caricamento dei modelli di riconoscimento di volto e voce associati all'utente, con relative soglie calcolate in fase di Tuning.
- Creazione di coppie volto/voce: ciascun elemento degli array dei volti sarà associato con tutti gli elementi dell'array delle voci.
- Determinazione dei valori di verità degli accoppiamenti volto/voce:se sia il volto che la voce sono esempi positivi per l'utente attivo, l'etichetta di verità avrà valore 1, avrà valore 0 altrimenti.
- Per ogni coppia volto/voce verrà fatta calcolare separatamente ai due modelli la predizione per l'elemento della coppia corrispondente.Basandosi su tali predizioni si deciderà se accettare o rigettare la coppia volto/voce.
- Se sia il modello per il riconoscimento del volto che quello per il riconoscimento della voce restituiranno una predizione positiva, allora la coppia volto/voce sarà classificata come positiva.
- Verrà infine calcolata la matrice di conclusione e l'accuratezza della combinazione.

Infine viene calcolata l'accuratezza media ottenuta su tutti gli utenti.

# 5. Conclusioni

Durante lo sviluppo si è fin da subito notato un miglioramento di prestazioni durante la combinazione dei sistemi.Tuttavia una volta terminato il progetto si è ritenuto opportuno eseguire delle misurazioni ufficiali di cui tener traccia.

Si è prima valutata in maniera separata l'accuratezza media dei due sistemi indipendenti,successivamente si è valutata l'accuratezza della combinazione,ottenendo come risultato un'aumento di accuratezza in tutti i test effettuati.

I modelli sono stati addestrati e testati per 5 volte ottenendo i seguenti risultati, da cui possiamo evincere che combinare due tecniche biometriche può portare ad avere un aumento delle prestazioni del sistema.

| # | Acc. Media volto | Acc. Media Voce | Acc. Media Combinazione |
|---|---|---|---|
| 1 | 0.89158 | 0.76597 | 0.96030 |
| 2 | 0.91464 | 0.73609 | 0.96565 |
| 3 | 0.94205 | 0.75632 | 0.97245 |
| 4 | 0.92959 | 0.71172 | 0.96301 |
| 5 | 0.92398 | 0.78344 | 0.97504 |