Lezione 8 - Indicizzazione e recupero dell’audio

https://www.youtube.com/watch?v=5Npa118hOH8

Classificazione dei file audio

Approccio generale

Si deve innanzitutto fare una macro-distinzione tra file:

- Parlato
- Musica
- Rumori

La gestione è diversificata per tipologie, e le query audio sono gestite su sottoinsiemi simili di brani audio

Proprietà principali dell’audio

I segnali audio sono rappresentati:

- Nel dominio temporale
- Nel dominio delle frequenze

Ciascun tipo di rappresentazione è idonea per l’estrazione di determinate caratteristiche

Time Domain

Tecnica più immediata e intuitiva per la rappresentazione di un sengale la cui ampiezza varia nel tempo. Il silenzio è rappresentato dallo Zero

Si assume che ogni campione audio sia rappresentato mediante un insieme di 16 bit

Energia media

> Indica la “rumorosità” del segnale audio

$$E = \frac{\sum_{n=0}^{N-1} x(n)^2}{N}$$

- E → Energia media
- N → Numero totale dei campioni valutati
- x(n) → Valore del camione n-esimo

Zero Crossing Rate

> Indica con quale frequenza l’ampiezza del segnale cambia di segno

$$ZCR = \frac{\sum_{n=1}^{N} |sgn[x(n)] - sgn[x(n-1)]|}{2N}$$

Dove:

$$sgn[x(n)] = \begin{cases} 1 \quad se \quad x(n) > 0 \\ -1 \quad se \quad s(n) < 0 \end{cases}$$

Silence Ratio

> Parametro che indica la proporzione di silenzio nel brano musicale. Periodo entro il quale i valori assoluti di ampiezza di un certo numero di campioni e per un certo tempo siano prossimi ad una soglia specifica

$$SR = \frac{S}{L}$$

- S → Somma totale dei periodi di silenzio
- L → Lunghezza totale del brano

Magnitudo medio

Viene introdotta poiché l'energia media aumenta troppo esponenzialmente per ampiezze troppo elevate

$$E = \frac{\sum_{n=0}^{N-1} |x(n)|}{N}$$

Dominio delle Frequenze

La rappresentazione nel dominio delle frequenze deriva dalla trasformazione del dominio temporale attraverso la trasformazione di Fourier

Bandwidth

> Indica la gamma (o range) delle frequenze

Armoniche

Un suono prodotto da un corpo vibrante non è mai pure, ma sarà sempre costituito da un insieme di frequenze multiple della frequenza di base

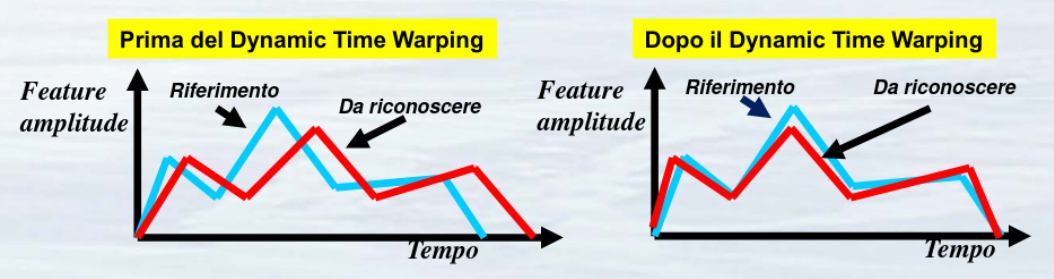
Metodi di Classificazione

Step by Step

1. Segnale audio in input
2. Valutazione del centroide
    a. Se alto → Musica
    b. Se basso → Potrebbe essere Voce e Musica
3. Valutazione del Silence Ratio
    a. Se basso → Musica
    b. Se alto → Potrebbe essere Voce o Musica
4. Valutazione del ZCR
    a. Se alto → Parlato
    b. Se basso → Musica

Time Warping

La tecnica del time warping cerca di “normalizzare” i frame dell’audio parlato, per far sì che coincidano con l’audio memorizzato sul sistema. Per ottenere questo risultato viene contratto o dilatato l’asse dei tempi, in modo da far coincidere picchi di segnale



Hidden Markov Model

Molto utilizzato per il riconoscimento sia dello scritto che del parlato

Impiega una rete sulla quale vengono definiti degli stati, una probabilità di transizione e una probabilità di generazione dei simboli

Reti Neurali Artificiali

Sono largamente impiegate per il riconoscimento e simulano i processi cognitivi del cervello umano

Si ha prima una fase di training, attraverso la quale verranno ottenuti vettori di caratteristiche per tarare i pesi dei link della rete

Nella fase di Recognition l’RNA seleziona il fenomena più verosimile basandosi sulle caratteristiche dei vettori