

## Approfondimento sperimentale – parte 1

Obiettivi di questa fase di progetto:

- prendere confidenza con la rappresentazione di gesti mediante serie temporali multivariate;
- utilizzo del modello vettoriale per la rappresentazione delle informazioni.

Trovate su moodle una cartella contenente i dati relativi a tre classi di gesti, di cui sono stati acquisiti dati spaziali da 20 sensori posizionati in diversi punti del corpo.

1. Scrivere un programma che crei un dizionario di “gesture words” a partire dal dataset fornito.

Dati una finestra di lunghezza  $w$  (parametro!), uno shift  $s$ , e una risoluzione  $r$ , si richiede che il programma

- legga tutti i file relativi ai gesti in una directory specificata
- normalizzi ciascun file sull'intervallo  $[-1,1]$
- discretizzi i valori di ciascuna serie mappandoli su  $2r$  valori diversi. Per far cio' si richiede di dividere l'intervallo  $[-1,1]$  in  $2r$  sotto-intervalli che siano “fasce gaussiane” di media 0 e deviazione standard 0,25);
- faccia scorrere la finestra  $w$  su ciascuna serie monovariata, shiftandola ad ogni passo di  $s$  unita', identificando in tal modo l'insieme  $D$  di *gesture words* nel dataset
- associ a ciascuna serie temporale univariata 3 “gesture vectors”, sulla base di
  - valore di TF
  - valore di TF-IDF
  - valore di TF-IDF2 (quest'ultimo calcolato considerando solo le parole relative a ciascun singolo gesto)

2. Scrivere un programma che consenta all'utente di

- selezionare un gesto  $d$  e visualizzarlo sotto forma di *heatmap* in scala di grigi, ed evidenziare (in base alla richiesta dell'utente)
  - le top-10 gesture words con valore di TF piu' alto;
  - le top-10 gesture words con valore di IDF piu' alto;
  - le top-10 gesture words con valore di IDF2 piu' alto;
  - le top-10 gesture words con valore di TF-IDF piu' alto;
  - le top-10 gesture words con valore di TF-IDF2 piu' alto.

3. Scrivere un programma che, dato un datafile  $d$ , identifichi le top-10 serie temporali multivariate piu' simili sulla base di

- sulla base dei valori TF;
- sulla base dei valori TF-IDF;
- sulla base dei valori TF-IDF2.