Hands-on sessions on Social Signal Processing

Mohamed CHETOUANI

mohamed.chetouani@sorbonne-universite.fr

TP1 : Intent recognition

Exercice 1. Automatic detection of speaker's intention from supra-segmental features

In this exercice, we consider a Human-Robot Interaction situation in which a Human is evaluating actions performed by the Kismet robot : approval or prohibition. The initial corpus contains a total of 1002 American English utterances of varying linguistic content produced by three female speakers in five classes of affective communicative intents. The classes are Approval, Attention, Prohibition Weak, Soothing, and Neutral utterances. The affective intents sound acted and are generally expressed rather strongly. The speech recordings are of variable length, mostly in the range of 1.8 - 3.25s.

We extracted prosodic features such as f_0 and energy. Files are respectively named $*.f_0$ and $*.en$ (time, value).

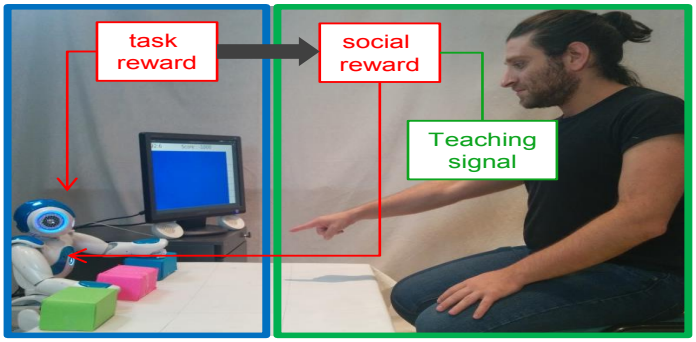


FIGURE 1 – Interactive Robot Learning : the human is continuously evaluating robot actions and providing feedbacks

The aim of this exercice is to develop a human feedback classifier : positive (approval) / negative (prohibition). This classifier might be used to teach robots and/or to guide robot's learning. Development of human feedback classifier :

1. Extraction of prosodic features (f_0 and energy)
2. Extraction of functionals (statistics) : mean, maximum, range, variance, median, first quartile, third quartile, mean absolute of local derivate
3. Check functionals for both voiced (i.e. $f_0 \neq 0$) and unvoiced segments. Which segments are suited for the approach ?
4. Build two databases by randomly extracting examples : Learning database (60%) and Test database
5. Train a classifer (k-NN, SVM or any classifier approach)
6. Evaluate and discuss the performance of the classifier

You will discuss the relevance of the parameters (f_0 et energy), the role of the functionals, the role of k, ratio of Learning/Test databases, random design of databases.

Exercice 2. Detection of multiple intents :

We consider the following intents : "Approval", "Prohibition" and "Attention"

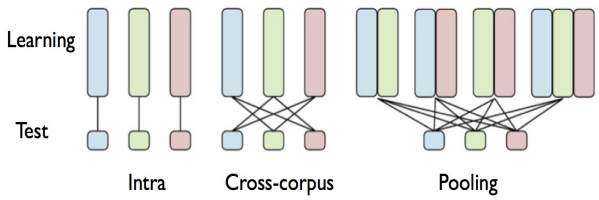
1. Extract the prosodic features (f_0 and energy) and their functionals
2. Develop a classifier for these three classses
3. Evaluate and discuss the performance of the classifier. We could use confusion matrices.

Exercice 3. Multi-corpus approach
Objective : Evaluate the impact of data on the performances

We consider both KISMET and BabyEars corpora. The multi-corpus approach will be applied only on the common intention classes.

1. Extract the prosodic features for Baby Ears (f_0 and energy).
2. Develop an utterance based classification with the k-NN method for BabyEars.

The multi-corpus approaches have been developed to enhance the performances of recognition systems by taking into account variabilities (speaker, annotation...). Several approaches have been proposed, we will focus on those described in the following figure :



For each of these conditions, compute the recognition rate. Discuss the interest of the combination of different situations (parent-infant, adult-robot) for designing robust systems.