INF642 : Socio-emotional Embodied Conversational Agents
Lab 3: Evaluation of the model
20/01/2020 – 1:30 pm

Objective: The objective of this lab is to evaluate the model of lab2.

Note: In this lab we are going to disregard Action Units 05, 06 and 07.

Model and Predictions

Download Lab 3 material from Moodle

Question 1: Inside "*train.py*", unpickle the following sets (*after # Question 1 comment*):

- *Training Set*

```
F0train, AU1_train, AU2_train, AU4_train, scalerAU1, scalerAU2, scalerAU4,
        ID_to_AU1, ID_to_AU2, ID_to_AU4, F0_to_ID, S1, S2, S4 =
                   numpy.load(…complete…,allow_pickle=True)
```

- *Validation Set*

```
F0validate, AU1_validate, AU2_validate, AU4_validate =
             numpy.load(…complete…,allow_pickle=True)
```

- *Test Set*

```
F0test, AU1_test, AU2_test, AU4_test =
      numpy.load(…complete…,allow_pickle=True)
```

For each set, print the shape of F0 and AU01. Comment.

Question 2: Compile the model. How many trainable parameters does the model have ? Comment.

Question 3: Train and fit the model.
(The training take some times, read the following questions while waiting)
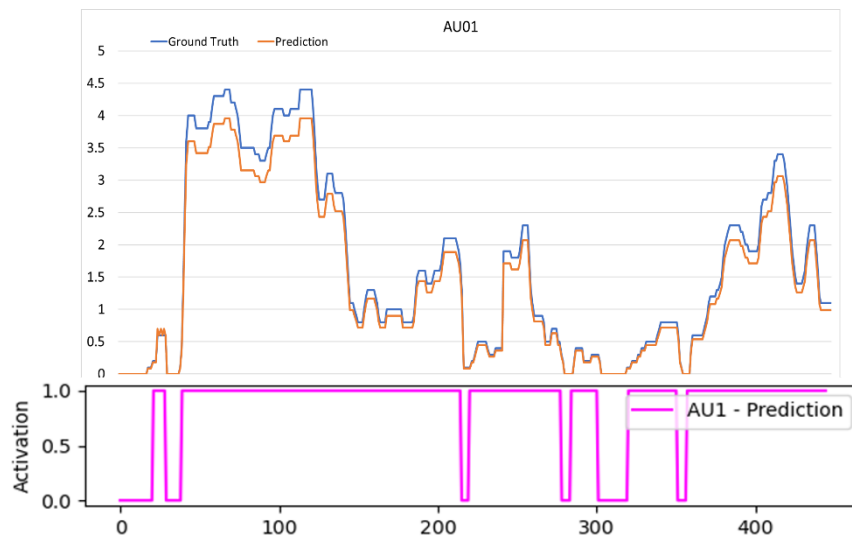
The function "*InferencePredictions*" runs the model in inference mode. It returns the predictions of AU01, AU02 and AU04 based on F0, for the first 15 words in the set.

Question 4: Use *"InferencePredictions"* to make predictions of Action Units for the first 15 words of the training set. **Save the predictions results in *AU01_15W_train*, *AU02_15W_train*, and *AU04_15W_train and groundtruths in AU01_15W_train_truth, AU02_15W_train_truth, and AU04_15W_train_truth. .**

Question 5: Use *"InferencePredictions"* to make predictions of Action Units for the first 15 words of the test set. **Save the predictions in results in *AU01_15W_test*, *AU02_15W_test*, and *AU04_15W_test* and groundtruths in *AU01_15W_test_truth*, *AU02_15W_test_truth*, and AU04_15W_test_truth.**

Objective Evaluation

We consider an AU as "***Activated***" (AU = 1) when its value is <u>greater than 0.5</u>, otherwise it is "***Not-Activated***" (AU = 0). The following figure depicts an example:

Question 6 : Compute the activation of your predictions on the 15 first words of the training and testing set and plot the activation of the first word of the testing set.

Question 7: For the training and test set, generate different plots of <u>activated action units</u> to illustrate predicted sequences Vs. their corresponding ground truth sequences, for each Action Unit.

<u>**Do NOT plot Start, End, and Padding tokens**</u>

Question 8 : To assess the quality of the generated gestures, apply the following measures on the test set:

- Root Mean Squared Error (RMSE)
- Pearson Correlation Coefficient (PCC)
- Activity Hit Ratio (AHR): the percentage of predicted AU activation with respect to the AU activation in ground truth.
- Non-Activity Hit Ratio (NAHR): the percentage of predicted AU non-activation with respect to AU non-activation in ground truth.

Question 9 : Plot and explain each of the metrics.

Question 10 : Comment your objective evaluation results.

Submission Details: submit your completed notebook with comments on moodle.