Project 3a

Harsh Agrawal
Department of Computer Science)
North Carolina State University
Raleigh, USA
hagrawa@ncsu.edu

Shraddha Dhayde
Department of Computer Science)
North Carolina State University
Raleigh, USA
sddhyade@ncsu.edu

Isha Pandya
Department of Computer Science)
North Carolina State University
Raleigh, USA
ipandya@ncsu.edu

I. Methodology

Before any classification model could be run to identify motion, the data of motions over gyro which was in sessions of 1-2 hours was prepossessed to get rolling average with window size 150. The model has been trained using Random Forest approach.

A. Extracted features:

First, we identified the features to be extracted from sessions data as per the Matlab script provided to us.
Features extracted are:

1) mean
2) covariance
3) skewness
4) kurtosis
5) maximum
6) minimum

B. Toolbox used:

We used the standard functions available in the pandas library to extract these functions. Namely, mean(), cov(), skewness(), kurtosis(), max(), min()

C. Approach used:

We then extracted the features on sliding window of 3 seconds (i.e. 150 rows of data) for four sets of data:

1) Arm data from Accelerometer device
2) Arm data from Gyroscope device
3) Wrist data from Accelerometer device
4) Wrist data from Gyroscope device

We then take the majority vote i.e. mode of the data from the rolling window results.

II. Model Training and Hyperparameter Selection

A. Model training

We used Random forest classifier model to train these extracted features. We split the training sessions data into training dataset (4 sessions) and validation dataset (2 sessions) and shuffled them in order to train the model.

B. Hyperparameter selection

We are using manual approach for hyper-parameter tuning. We have used the following set of hyper-parameters for tuning our Random Forest classifier model:

1) n_estimators = [10,15,20]
2) max_depth = [5,10,15]
3) criterion = ['gini', 'entropy']

The best hyperparameters obtained are n_estimators = 15, max_depth = 10, and criterion = 'entropy'. A Validation accuracy of 90.06% is obtained.

III. Evaluation

We performed evaluation on the validation set i.e. Session 01 and Session 13 and observed the following results:

We have plotted the results for prediction.txt and detection.txt for respective sessions.

We got an accuracy of 89.56% for Session 1 and 84.59% for Session 13.

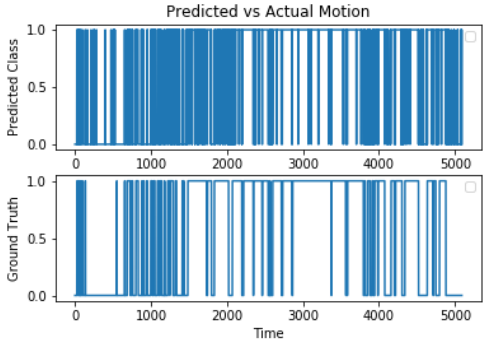

Fig. 1. Validation dataset overlap for session 1

The results obtained from evaluation script are as follows:
Average over Session

1) Recall: 0.9577171696443458
2) Precision: 0.7068447291208215
3) Accuracy: 0.8642153679734542
4) F1: 0.7936369462751904

Entire Data

1) Recall: 0.9170811456597604
2) Precision: 0.8836594971220842
3) Accuracy: 0.8814028509599012

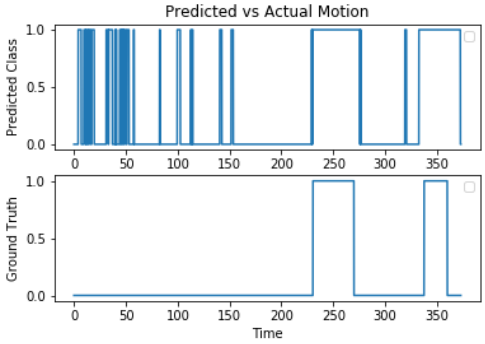

Fig. 2. Validation dataset overlap for session 13

4) F1: 0.9000601694000031