Activity Recognition

Assignment #1

Vanessa Ulloa | CSE 572: Data Mining | 07 Apr 2019

Table of Contents

[1 Phase 1: Data Collection 2](#_Toc5481519)

[Phase 2: Feature Extraction 4](#_Toc5481520)

[Fast Fourier Transformation 5](#_Toc5481521)

[Explanation 5](#_Toc5481522)

[Intiuition 5](#_Toc5481523)

[Matlab code 5](#_Toc5481524)

[Graphs 5](#_Toc5481525)

[Discuss 5](#_Toc5481526)

[Standard Deviation 5](#_Toc5481527)

[Explanation 5](#_Toc5481528)

[Intiuition 5](#_Toc5481529)

[Matlab code 5](#_Toc5481530)

[Graphs 5](#_Toc5481531)

[Discuss 5](#_Toc5481532)

[Minimum 5](#_Toc5481533)

[Explanation 5](#_Toc5481534)

[Intiuition 5](#_Toc5481535)

[Matlab code 5](#_Toc5481536)

[Graphs 5](#_Toc5481537)

[Discuss 5](#_Toc5481538)

[Maximum 5](#_Toc5481539)

[Explanation 5](#_Toc5481540)

[Intiuition 5](#_Toc5481541)

[Matlab code 5](#_Toc5481542)

[Graphs 5](#_Toc5481543)

[Discuss 5](#_Toc5481544)

[Average 5](#_Toc5481545)

[Explanation 5](#_Toc5481546)

[Intiuition 5](#_Toc5481547)

[Matlab code 5](#_Toc5481548)

[Graphs 5](#_Toc5481549)

[Discuss 5](#_Toc5481550)

[Phase 3: Feature Selection 6](#_Toc5481551)

[Arranging the Feature Matrix 6](#_Toc5481552)

[Execution of PCA 6](#_Toc5481553)

[PCA eigenvectors 6](#_Toc5481554)

[PCA results 6](#_Toc5481555)

[PCA comments 6](#_Toc5481556)

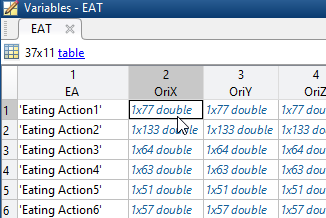
# Phase 1: Data Collection

The first part of this project was to organize the raw data. The files that were given per user were one file with IMU sensor data, EMG sensor data and a file with frame numbers for either an action with a fork or spoon for a user. The file with the frame numbers (labeled {userid}.txt) contains 3 columns of data, the first column is a start frame number and the second column is the end frame number; the third column is not used. The significance of these start and end frames are that they are in relation to video taken of a user with the Myo wrist band and when they are engaging in an eating action. We will use this frame numbers in order to synchronize the IMU and EMG data and sort out which data belongs to eating actions and which data belongs to non-eating actions.

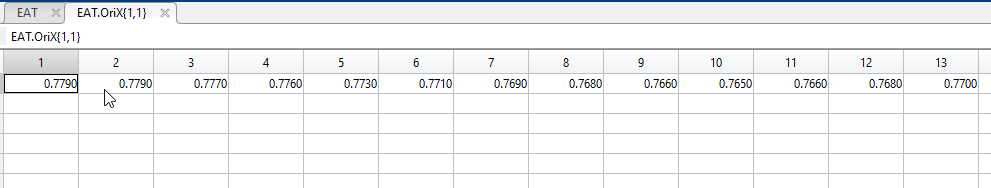
The video data will be the ground truth data used in this project. The assumed frames per second in this video is 30 fps. The sampling rate for the IMU data is 50 Hz or 50 samples per second, the sampling rate for the EMG data is 200 Hz or 200 samples per second. In order to synchronize the IMU data the start frame in the ground truth file was multiplied by (50/30) and similarly the end frame. This new start and end sampling value is used in order to retrieve the start and end row from the IMU file that correlates to the eating action frames. The same method is used for the EMG data except for the calculation is (200/30) since the sampling rate is not the same as the IMU sampling rate. Similarly, the new start and end sampling value is used to retrieve the data from the EMG file that correlates to the eating action.

The non-eating data was found using a similar method except that we use values in between a start frame and the previous end frame in order to indicate non-eating actions. For example, we iterate through the ground truth file start at *i*=2, then we apply the calculation (either for EMG or IMU) to the start frame at row *i* and the end frame at row *i-1* (the previous row). This range gives us the rows that correlate in either the IMU or EMG files for a non-eating action.

The IMU data that is provided includes sensor data for orientation, accelerometer and gyroscope. This results in 10 different sensor data points (OriX, OriY, OriZ, OriW, AccX, AccY, AccZ, GyroX, GyroY, GyroZ). The EMG data that is provided includes 8 sensor data points (EMG1-8). This data is stored into a table of cell arrays similar to below:



The value EAT.OriX{1,1} returns the value:



All the values for Orientation X during Eating Action 1.

TODO: something about CSV files.

# Phase 2: Feature Extraction

The data is organized into a table of cell arrays where each table row is a unique eating action and each column is a sensor. The cell value contains the raw data for that eating action and that sensor as a time series. The sensor data varies based on action and sensor and is not of consistent length with the other cells. The feature extractions chosen were:

1. FFT (Fast Fourier Transformation)
2. Standard Deviation
3. Minimum
4. Maximum
5. Average

## Fast Fourier Transformation

### Explanation

### Intiuition

FFT is used to change the raw data from a time domain to a frequency domain so that we can see how fast the data is changing.

### Matlab code

### Graphs

### Discuss

## Standard Deviation

### Explanation

### Intiuition

### Matlab code

### Graphs

### Discuss

## Minimum

### Explanation

### Intiuition

### Matlab code

### Graphs

### Discuss

## Maximum

### Explanation

### Intiuition

### Matlab code

### Graphs

### Discuss

## Average

### Explanation

### Intiuition

### Matlab code

### Graphs

### Discuss

# Phase 3: Feature Selection

## Arranging the Feature Matrix

## Execution of PCA

## PCA eigenvectors

## PCA results

## PCA comments