

Codebook

Esko Nuutila

14 Jan 2017

This codebook is based on the original file `features_info.txt` of the Human Activity Recognition Using Smartphones Dataset. I have edited the feature list to contain only the variables that are available in my tidy dataset.

The tidy dataset summarizes the original dataset by combining the train and test datasets, and grouping and averaging the features for each subject (30 people) and each type of activity (“WALKING”, “WALKING_UPSTAIRS”, “WALKING_DOWNSTAIRS”, “SITTING”, “STANDING”, “LAYING”). For each pair of subject and activity, the tidy dataset has 66 variables that are a subset of the variables in the original dataset.

The features in the original data sets are described in file `features_info.txt` as follows:

The features selected for this database come from the accelerometer and gyroscope 3-axial raw signals `tAcc-XYZ` and `tGyro-XYZ`. These time domain signals (prefix 't' to denote time) were captured at a constant rate of 50 Hz. Then they were filtered using a median filter and a 3rd order low pass Butterworth filter with a corner frequency of 20 Hz to remove noise. Similarly, the acceleration signal was then separated into body and gravity acceleration signals (`tBodyAcc-XYZ` and `tGravityAcc-XYZ`) using another low pass Butterworth filter with a corner frequency of 0.3 Hz.

Subsequently, the body linear acceleration and angular velocity were derived in time to obtain Jerk signals (`tBodyAccJerk-XYZ` and `tBodyGyroJerk-XYZ`). Also the magnitude of these three-dimensional signals were calculated using the Euclidean norm (`tBodyAccMag`, `tGravityAccMag`, `tBodyAccJerkMag`, `tBodyGyroMag`, `tBodyGyroJerkMag`).

Finally a Fast Fourier Transform (FFT) was applied to some of these signals producing `fBodyAcc-XYZ`, `fBodyAccJerk-XYZ`, `fBodyGyro-XYZ`, `fBodyAccJerkMag`, `fBodyGyroMag`, `fBodyGyroJerkMag`. (Note the 'f' to indicate frequency domain signals).

These signals were used to estimate variables of the feature vector for each pattern: '-XYZ' is used to denote 3-axial signals in the X, Y and Z directions.

`tBodyAcc-XYZ`
`tGravityAcc-XYZ`
`tBodyAccJerk-XYZ`
`tBodyGyro-XYZ`
`tBodyGyroJerk-XYZ`
`tBodyAccMag`
`tGravityAccMag`
`tBodyAccJerkMag`
`tBodyGyroMag`
`tBodyGyroJerkMag`
`fBodyAcc-XYZ`
`fBodyAccJerk-XYZ`
`fBodyGyro-XYZ`
`fBodyAccMag`
`fBodyAccJerkMag`
`fBodyGyroMag`

fBodyGyroJerkMag

In the tidy data set, each pattern `pattern-XYZ` above is represented by six variables: `pattern-mean()-X`, `pattern-std()-X`, `pattern-mean()-Y`, `pattern-std()-Y`, `pattern-mean()-Z`, `pattern-std()-Z`. Each pattern without the XYZ ending is represented by two variables: `pattern-mean()` and `pattern-std()`.

The values for these variables are computed by combining the train and test datasets, selecting the columns of the variables from the combined dataset, grouping by `activity` and `subject` and taking the mean for each variable. See file `run_analysis.R` for details.

In addition the dataset contains variables `activity` and `subject`.

The complete list of variables (68) in the tidy dataset is:

- `activity`
- `subject`
- `tBodyAcc-mean()-X`
- `tBodyAcc-mean()-Y`
- `tBodyAcc-mean()-Z`
- `tBodyAcc-std()-X`
- `tBodyAcc-std()-Y`
- `tBodyAcc-std()-Z`
- `tGravityAcc-mean()-X`
- `tGravityAcc-mean()-Y`
- `tGravityAcc-mean()-Z`
- `tGravityAcc-std()-X`
- `tGravityAcc-std()-Y`
- `tGravityAcc-std()-Z`
- `tBodyAccJerk-mean()-X`
- `tBodyAccJerk-mean()-Y`
- `tBodyAccJerk-mean()-Z`
- `tBodyAccJerk-std()-X`
- `tBodyAccJerk-std()-Y`
- `tBodyAccJerk-std()-Z`
- `tBodyGyro-mean()-X`
- `tBodyGyro-mean()-Y`
- `tBodyGyro-mean()-Z`
- `tBodyGyro-std()-X`
- `tBodyGyro-std()-Y`
- `tBodyGyro-std()-Z`
- `tBodyGyroJerk-mean()-X`
- `tBodyGyroJerk-mean()-Y`
- `tBodyGyroJerk-mean()-Z`
- `tBodyGyroJerk-std()-X`
- `tBodyGyroJerk-std()-Y`
- `tBodyGyroJerk-std()-Z`
- `tBodyAccMag-mean()`
- `tBodyAccMag-std()`
- `tGravityAccMag-mean()`
- `tGravityAccMag-std()`
- `tBodyAccJerkMag-mean()`
- `tBodyAccJerkMag-std()`
- `tBodyGyroMag-mean()`
- `tBodyGyroMag-std()`
- `tBodyGyroJerkMag-mean()`

- tBodyGyroJerkMag-std()
- fBodyAcc-mean()-X
- fBodyAcc-mean()-Y
- fBodyAcc-mean()-Z
- fBodyAcc-std()-X
- fBodyAcc-std()-Y
- fBodyAcc-std()-Z
- fBodyAccJerk-mean()-X
- fBodyAccJerk-mean()-Y
- fBodyAccJerk-mean()-Z
- fBodyAccJerk-std()-X
- fBodyAccJerk-std()-Y
- fBodyAccJerk-std()-Z
- fBodyGyro-mean()-X
- fBodyGyro-mean()-Y
- fBodyGyro-mean()-Z
- fBodyGyro-std()-X
- fBodyGyro-std()-Y
- fBodyGyro-std()-Z
- fBodyAccMag-mean()
- fBodyAccMag-std()
- fBodyBodyAccJerkMag-mean()
- fBodyBodyAccJerkMag-std()
- fBodyBodyGyroMag-mean()
- fBodyBodyGyroMag-std()
- fBodyBodyGyroJerkMag-mean()
- fBodyBodyGyroJerkMag-std()