Codebook Tidy Dataset

Volha Leusha

11.15.2018

Project Description

Tidy Dataset, Created for "Getting and Cleaning Data" Coursera Course, Assignment 4.

Collection of the raw data

This Tidy dataset is based on Human Activity Recognition Using Smartphones Dataset Version 1.0. The full describtion can be obtained here: links

Notes on the original (raw) data

The experiments have been carried out with a group of 30 volunteers within an age bracket of 19-48 years. Each person performed six activities (WALKING, WALKING_UPSTAIRS, WALKING_DOWNSTAIRS, SITTING, STANDING, LAYING) wearing a smartphone (Samsung Galaxy S II) on the waist. Using its embedded accelerometer and gyroscope, we captured 3-axial linear acceleration and 3-axial angular velocity at a constant rate of 50Hz. The experiments have been video-recorded to label the data manually. The obtained dataset has been randomly partitioned into two sets, where 70% of the volunteers was selected for generating the training data and 30% the test data.

The sensor signals (accelerometer and gyroscope) were pre-processed by applying noise filters and then sampled in fixed-width sliding windows of 2.56 sec and 50% overlap (128 readings/window). The sensor acceleration signal, which has gravitational and body motion components, was separated using a Butterworth low-pass filter into body acceleration and gravity. The gravitational force is assumed to have only low frequency components, therefore a filter with 0.3 Hz cutoff frequency was used. From each window, a vector of features was obtained by calculating variables from the time and frequency domain.

Creating the tidy datafile

Guide to create the tidy data file

1. Download the data from:
   links
2. Unzip data to R directory folder

3. Open run_analysis.R

4. In case if additional libraries are not installes, install them: data.table, dplyr, plyr

5. Run run_analysis.R

Cleaning of the data

Tidy dataset was creaded from original dataset, performing following steps:

1. Merged the training and the test sets to create one data set.

2. Extracted only the measurements on the mean and standard deviation for each measurement.

3. Used descriptive activity names to name the activities in the data set.

4. Appropriately labeled the data set with descriptive variable names.

5. From the data set in step 4, created a second, independent tidy data set with the average of each variable for each activity and each subject.

For more details about refer to readme.txt.

Description of the variables in the tiny_data.txt file

General description of the file including:

- Dimensions of the dataset: 180 rows x 88 columns
- Summary of the data:Dataset is grouped by 2 variables: "y"(1st column)- 6 possible and "subject"(2nd column). There are 6 possible lables "y" and 30 possible "subjects", therefore there are 6x30=180 rows in dataset. For each group "y"+"subject" there is one possible value of "feature variables" (3:88 columns). This value is obtained by taking colmean() from original dataset.
- Variables present in the dataset:

  "label" (column 1) activity labels
  "subject" (column 2) : identifier of the subject who carried out the experiment
  "feature vector" (column 3:88): mean and std measurements of time and frequency domain variables

label

Activity label.

- Class: factor

- Unique values:
  LAYING
  SITTING
  STANDING
  WALKING
  WALKING_DOWNSTAIRS
  WALKING_UPSTAIRS

- Unit of measurement: NA

subject

Identifier of the subject who carried out the experiment.

- Class: int
- Unique values: 1-30
- Unit of measurement: NA

Feature Vector

Vector containing mean and std measurements of time and frequency domain variables. Columns 3:88 of dataset.

Contains following variables:
3 TimeBodyAccMeanX
4 TimeBodyAccMeanY
5 TimeBodyAccMeanZ
6 TimeBodyAccStdX
7 TimeBodyAccStdY
8 TimeBodyAccStdZ
9 TimeGravityAccMeanX
10 TimeGravityAccMeanY
11 TimeGravityAccMeanZ
12 TimeGravityAccStdX
13 TimeGravityAccStdY
14 TimeGravityAccStdZ
15 TimeBodyAccJerkMeanX
16 TimeBodyAccJerkMeanY
17 TimeBodyAccJerkMeanZ
18 TimeBodyAccJerkStdX
19 TimeBodyAccJerkStdY
20 TimeBodyAccJerkStdZ
21 TimeBodyGyroMeanX
22 TimeBodyGyroMeanY
23 TimeBodyGyroMeanZ
24 TimeBodyGyroStdX
25 TimeBodyGyroStdY
26 TimeBodyGyroStdZ
27 TimeBodyGyroJerkMeanX
28 TimeBodyGyroJerkMeanY
29 TimeBodyGyroJerkMeanZ
30 TimeBodyGyroJerkStdX
31 TimeBodyGyroJerkStdY
32 TimeBodyGyroJerkStdZ
33 TimeBodyAccMagMean
34 TimeBodyAccMagStd
35 TimeGravityAccMagMean
36 TimeGravityAccMagStd
37 TimeBodyAccJerkMagMean
38 TimeBodyAccJerkMagStd
39 TimeBodyGyroMagMean
40 TimeBodyGyroMagStd
41 TimeBodyGyroJerkMagMean
42 TimeBodyGyroJerkMagStd
43 FrequencyBodyAccMeanX
44 FrequencyBodyAccMeanY
45 FrequencyBodyAccMeanZ
46 FrequencyBodyAccStdX
47 FrequencyBodyAccStdY
48 FrequencyBodyAccStdZ
49 FrequencyBodyAccMeanX
50 FrequencyBodyAccMeanY

51 FrequencyBodyAccMeanZ
52 FrequencyBodyAccJerkMeanX
53 FrequencyBodyAccJerkMeanY
54 FrequencyBodyAccJerkMeanZ
55 FrequencyBodyAccJerkStdX
56 FrequencyBodyAccJerkStdY
57 FrequencyBodyAccJerkStdZ
58 FrequencyBodyAccJerkMeanX
59 FrequencyBodyAccJerkMeanY
60 FrequencyBodyAccJerkMeanZ
61 FrequencyBodyGyroMeanX
62 FrequencyBodyGyroMeanY
63 FrequencyBodyGyroMeanZ
64 FrequencyBodyGyroStdX
65 FrequencyBodyGyroStdY
66 FrequencyBodyGyroStdZ
67 FrequencyBodyGyroMeanX
68 FrequencyBodyGyroMeanY
69 FrequencyBodyGyroMeanZ
70 FrequencyBodyAccMagMean
71 FrequencyBodyAccMagStd
72 FrequencyBodyAccMagMean
73 FrequencyBodyBodyAccJerkMagMean
74 FrequencyBodyBodyAccJerkMagStd
75 FrequencyBodyBodyAccJerkMagMean
76 FrequencyBodyBodyGyroMagMean
77 FrequencyBodyBodyGyroMagStd
78 FrequencyBodyBodyGyroMagMean
79 FrequencyBodyBodyGyroJerkMagMean
80 FrequencyBodyBodyGyroJerkMagStd
81 FrequencyBodyBodyGyroJerkMagMean
82 Angle(TimeBodyAccMean,Gravity)
83 Angle(TimeBodyAccJerkMean),GravityMean)
84 Angle(TimeBodyGyroMean,GravityMean)
85 Angle(TimeBodyGyroJerkMean,GravityMean)
86 Angle(X,GravityMean)
87 Angle(Y,GravityMean)
88 Angle(Z,GravityMean)

- Class: numeric
- Unique values: [-1:1]
- Unit of measurement: NA
- Names Descriptor:
  Mean: Mean value if the signal
  Std: Standard Deviation value of the signal
  Y: Y-axis signal
  X: X-axis signal
  Z: Z-axis signal
  Acc: captured by accelerometer
  Gyro: captured by gyroscope

Time: A time domain signal, captured at a constant rate of 50 Hz. Filtered using a median filter and a 3rd order low pass Butterworth filter with a corner frequency of 20 Hz to remove noise
Frequency: Fast Fourier Transform (FFT) is applied on signal and shifts it to frequency domain
Angle: Averaged in signal window sample
Gravity: Low pass Butterworth filter with a corner frequency of 0.3 Hz separated it into gravity signal
Body: Low pass Butterworth filter with a corner frequency of 0.3 Hz separated it into body signal
Mag: The magnitude of these three-dimensional signals were calculated using the Euclidean norm
Jerk: The body linear acceleration and angular velocity were derived in time to obtain signal