Codebook for Class Project "Getting and Cleaning Data" Coursera Course

Ravi Mathi

This codebook describes a piece of R Code used to prepare tidy data that can be used for later analysis.

Introduction

One of the most exciting areas in all of data science right now is wearable computing - see for example this article . Companies like Fitbit, Nike, and Jawbone Up are racing to develop the most advanced algorithms to attract new users. The data linked to from the course website represent data collected from the accelerometers from the Samsung Galaxy S smartphone. A full description is available at the site where the data was obtained:

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.

http://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones

Variables description

test_tabl - read the test data set

train_tabl - read in the train data set

sub_features_Col - data frame with descriptive column names from file Subset_features_Descriptive.txt derived from the provided features_info.txt

sub_train_tabl - holds the subset of the training data

sub_test_tabl - - holds the subset of the testing data

trainAndtest_Tabl - dataframe of combined test and training data

y_trainVector - read in the training activity labels from y_train.txt

y_testVector - read in the testing activity labels from y_test.txt

combinedActivity - vector of combined activity codes

activityLabelsTabl - reads in the 6 activity labels from activity_labels.txt

ActLen - number of rows in combinedActivity

trainSubjVector - read in the training subjects vector from subject_train.txt

testSubjVector - read in the test subjects vector from subject_test.txt

combinedSubjectList - combines the above two vectors

TidyDataSetdf - Holds the final tidy data set

The 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.

For each record in the dataset it is provided:

- Triaxial acceleration from the accelerometer (total acceleration) and the estimated body acceleration.

- Triaxial Angular velocity from the gyroscope.

- A 561-feature vector with time and frequency domain variables.

- Its activity label.

- An identifier of the subject who carried out the experiment.

Description of each column in the Tidy data set

1, "(T)Body Acceleration Mean X-Axis"

2, "(T)Body Acceleration Mean Y-Axis"

3, "(T)Body Acceleration Mean Z-Axis"

4, "(T)Body Acceleration Standard Deviation X-Axis"

5, "(T)Body Acceleration Standard Deviation Y-Axis"

6, "(T)Body Acceleration Standard Deviation Z-Axis"

41, "(T)Gravity Acceleration Mean X-Axis"

42, "(T)Gravity Acceleration Mean Y-Axis"

43, "(T)Gravity Acceleration Mean Z-Axis"

44, "(T)Gravity Acceleration Standard Deviation X-Axis"

45, "(T)Gravity Acceleration Standard Deviation Y-Axis"

46, "(T)Gravity Acceleration Standard Deviation Z-Axis"

81, "(T)Body Acceleration Jerk Mean X-Axis"

82, "(T)Body Acceleration Jerk Mean Y-Axis"

83, "(T)Body Acceleration Jerk Mean Z-Axis"

84, "(T)Body Acceleration Jerk Standard Deviation X-Axis"

85, "(T)Body Acceleration Jerk Standard Deviation Y-Axis"

86, "(T)Body Acceleration Jerk Standard Deviation Z-Axis"

121, "(T)Body Gyroscope Mean X-Axis"

122, "(T)Body Gyroscope Mean Y-Axis"

123, "(T)Body Gyroscope Mean Z-Axis"

124, "(T)Body Gyroscope Standard Deviation X-Axis"

125, "(T)Body Gyroscope Standard Deviation Y-Axis"

126, "(T)Body Gyroscope Standard Deviation Z-Axis"

161, "(T)Body Gyroscope Jerk Mean X-Axis"

162, "(T)Body Gyroscope Jerk Mean Y-Axis"

163, "(T)Body Gyroscope Jerk Mean Z-Axis"

164, "(T)Body Gyroscope Jerk Standard Deviation X-Axis"

165, "(T)Body Gyroscope Jerk Standard Deviation Y-Axis"

166, "(T)Body Gyroscope Jerk Standard Deviation Z-Axis"

201, "(T)Body Acceleration Magnitude Mean"

202, "(T)Body Acceleration Magnitude Standard Deviation"

214, "(T)Gravity Acceleration Magnitude Mean"

215, "(T)Gravity Acceleration Magnitude Standard Deviation"

227, "(T)Body AccelerationJerk Magnitude Mean"

228, "(T)Body Acceleration Jerk Magnitude Standard Deviation"

240, "(T)Body GyroscopeMagnitude Mean"

241, "(T)Body GyroscopeMagnitude Standard Deviation"

253, "(T)Body Gyroscope Jerk Magnitude Mean"

254, "(T)Body Gyroscope Jerk Magnitude Standard Deviation"

266, "(F)Body Acceleration Mean X-Axis"

267, "(F)Body Acceleration Mean Y-Axis"

268, "(F)Body Acceleration Mean Z-Axis"

269, "(F)Body Acceleration Standard Deviation X-Axis"

270, "(F)Body Acceleration Standard Deviation Y-Axis"

271, "(F)Body Acceleration Standard Deviation Z-Axis"

345, "(F)Body Acceleration Jerk Mean X-Axis"

346, "(F)Body Acceleration Jerk Mean Y-Axis"

347, "(F)Body Acceleration Jerk Mean Z-Axis"

348, "(F)Body Acceleration Jerk Standard Deviation X-Axis"

349, "(F)Body Acceleration Jerk Standard Deviation Y-Axis"

350, "(F)Body Acceleration Jerk Standard Deviation Z-Axis"

373, "(F)Body Acceleration Jerk MeanFreq() X-Axis"

374, "(F)Body Acceleration Jerk MeanFreq() Y-Axis"

424, "(F)Body Gyroscope Mean X-Axis"

425, "(F)Body Gyroscope Mean Y-Axis"

426, "(F)Body Gyroscope Mean Z-Axis"

427, "(F)Body Gyroscope Standard Deviation X-Axis"

428, "(F)Body Gyroscope Standard Deviation Y-Axis"

429, "(F)Body Gyroscope Standard Deviation Z-Axis"

503, "(F)Body Acceleration Magnitude Mean"

504, "(F)Body Acceleration Magnitude Standard Deviation"

529, "(F)Body-Body Gyroscope Magnitude Mean"

530, "(F)Body-Body Gyroscope Magnitude Standard Deviation"

542, "(F)Body-Body Gyroscope Jerk Magnitude Mean"

543, "(F)Body-BodyGyroscope Jerk Magnitude Standard Deviation"

Transformations performed to clean up the data

The code has comments all along describing the transformations performed; Overall the following transformations were done:shows the

- Merged the training and the test sets to create one data set.
- Extracted only the measurements on the mean and standard deviation for each measurement.
- Provided descriptive activity names to name the activities in the data set
- Provided descriptive variable names for the data set
- And finally created a second, independent tidy data set with the average of each variable

Tidy data results snapshot.

This shows how the dity data set loks when viewed in the RStudio interface

Subject	Activity	(T)Body Acceleration Mean X-Axis	(T)Body Acceleration Mean Y-Axis	(T)Body Acceleration Mean Z-Axis	(T)Body Accel
1	LAYING	0.2215982	-0.040513953	-0.11320355	-0.928056469
2	LAYING	0.2813734	-0.018158740	-0.10724561	-0.974059465
3	LAYING	0.2755169	-0.018955679	-0.10130048	-0.982776639
4	LAYING	0.2635592	-0.015003184	-0.11068815	-0.954193739
5	LAYING	0.2783343	-0.018304212	-0.10793760	-0.965934510
6	LAYING	0.2486565	-0.010252917	-0.13311957	-0.934049422
7	LAYING	0.2501767	-0.020441152	-0.10136104	-0.936513610
8	LAYING	0.2612543	-0.021228173	-0.10224537	-0.943041215
9	LAYING	0.2591955	-0.020526822	-0.10754972	-0.942333090
10	LAYING	0.2802306	-0.024294484	-0.11716864	-0.968283685
11	LAYING	0.2805930	-0.017659805	-0.10878658	-0.984777330
12	LAYING	0.2601134	-0.017520392	-0.10816013	-0.955318680
13	LAYING	0.2767164	-0.020440454	-0.10433186	-0.968891989
14	LAYING	0.2332754	-0.011342465	-0.08683333	-0.917501887
15	LAYING	0.2894757	-0.016629654	-0.11853024	-0.972255552
16	LAYING	0.2742272	-0.016610351	-0.10731049	-0.973691429
17	LAYING	0.2697801	-0.016846201	-0.10700628	-0.972960567
18	LAYING	0.2746916	-0.017393768	-0.10769893	-0.984527649