

Getting and Cleaning Data Project Readme File

MARIANO MOLINA GARCIA

09/12/2019

Introduction

The purpose of this project is to demonstrate your ability to collect, work with, and clean a data set. The goal is to prepare tidy data that can be used for later analysis. The goals will include:

- 1) a tidy data set as described below,
- 2) a link to a Github repository with your script for performing the analysis.
- 3) a code book that describes the variables, the data, and any transformations or work performed to clean up the data called CodeBook.md.
- 4) Include a README.md file (this file) in the repo with the scripts. In this file it will be explained how all of the scripts work and how they are connected.

Original Data Info

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:

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

Here are the data for the project:

<https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset.zip>

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. See 'features_info.txt' for more details.

For each record 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.

The dataset includes the following files:

- 'README.txt'
- 'features_info.txt': Shows information about the variables used on the feature vector.
- 'features.txt': List of all features.
- 'activity_labels.txt': Links the class labels with their activity name.
- 'train/X_train.txt': Training set.
- 'train/y_train.txt': Training labels.
- 'test/X_test.txt': Test set.
- 'test/y_test.txt': Test labels.

The following files are available for the train and test data. Their descriptions are equivalent.

- 'train/subject_train.txt': Each row identifies the subject who performed the activity for each window sample. Its range is from 1 to 30.
- 'train/Inertial Signals/total_acc_x_train.txt': The acceleration signal from the smartphone accelerometer X axis in standard gravity units 'g'. Every row shows a 128 element vector. The same description applies for the 'total_acc_x_train.txt' and 'total_acc_z_train.txt' files for the Y and Z axis.
- 'train/Inertial Signals/body_acc_x_train.txt': The body acceleration signal obtained by subtracting the gravity from the total acceleration.
- 'train/Inertial Signals/body_gyro_x_train.txt': The angular velocity vector measured by the gyroscope for each window sample. The units are radians/second.

Files included in the Repo

1) Script run_analysis.R

This script include all the steps to get and clean the data. The files to be processed to create the tidy data set are:

The steps implemented are: a.Merges the training and the test sets to create one data set. b.Extracts only the measurements on the mean and standard deviation for each measurement. c.Uses descriptive activity names to name the activities in the data set d.Appropriately labels the data set with descriptive variable names. e.From the data set in step 4, creates a second, independent tidy data set with the average of each variable for each activity and each subject. f.Generate the structure info and the summary info of the different columns in the tidy data set.

2) File TidyDataset.txt

Txt file with the clean data set.

3) File TidyDatasetStr.txt

Txt file with the structure info of the columns in the clean data set.

4) File TidyDatasetSummary.txt

Txt file with the summary info of the columns in the clean data set.

5) Readme.md file

This file with general info about the project.

6) Codebook.md file

File with info of the process carried out to create the tidy data and info about the data included in the final tidy data set.