

Getting and Cleaning Data Course Project

Description

Additional information about the variables, data and transformations used in the course project for the Johns Hopkins Getting and Cleaning Data course.

Source Data

A full description of the data used in this project can be found at [The UCI Machine Learning Repository](#)

[The source data for this project can be found here.](#)

Data Set Information

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.

Getting and Cleaning Data Course Project

Attribute Information

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.

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

After setting the source directory for the files, read into tables the data located in

- features.txt
- activity_labels.txt
- subject_train.txt
- x_train.txt
- y_train.txt
- subject_test.txt
- x_test.txt
- y_test.txt

Assign column names and merge to create one data set.

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

Create vector for defining ID, mean and standard deviation. Subset this data to keep only the necessary columns.

Getting and Cleaning Data Course Project

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

Merge data subset with the activityLabels table to include the descriptive activity names

Section 4. Appropriately label the data set with descriptive activity names.

merg_train, merg_test merged_sets, merged_setMean_Std

Section 5. Create a second, independent tidy data set with the average of each variable for each activity and each subject.

Produced a data set with the average of each variable for each activity and subject.