Data Dictionary - Tidy Version of Mean and Standard Deviation Measurememnts from Smartphone-Based Recognition of Human Activities and Postural Transitions Data Set

Josh Kimrey

11/10/2020

original data was obtained from: http://archive.ics.uci.edu/ml/datasets/Smartphone-Based+Recognition+of+Human+Activities+and+Postural+Transitions

(description from site)

Data Set Information:

The experiments were carried out with a group of 30 volunteers within an age bracket of 19-48 years. They performed a protocol of activities composed of six basic activities: three static postures (standing, sitting, lying) and three dynamic activities (walking, walking downstairs and walking upstairs). The experiment also included postural transitions that occurred between the static postures. These are: stand-to-sit, sit-to-stand, sit-to-lie, lie-to-sit, stand-to-lie, and lie-to-stand. All the participants were wearing a smartphone (Samsung Galaxy S II) on the waist during the experiment execution. We captured 3-axial linear acceleration and 3-axial angular velocity at a constant rate of 50Hz using the embedded accelerometer and gyroscope of the device. The experiments were video-recorded to label the data manually. The obtained dataset was 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 561 features was obtained by calculating variables from the time and frequency domain. See 'features_info.txt' for more details.

Attribute Information:

The dataset is then divided in two parts and they can be used separately.

1. Inertial sensor data

- Raw triaxial signals from the accelerometer and gyroscope of all the trials with with participants.
- The labels of all the performed activities.

2. Records of activity windows. Each one composed of:

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

- Its associated activity label.
- An identifier of the subject who carried out the experiment.

Notes

- Features (excluding "id" and "activity") are normalized and bounded within [-1,1].
- The units used for the accelerations (total and body) are 'g's (gravity of earth -> 9.80665 m/seg2).
- The gyroscope units are rad/seg.

(end description from site)

Brief Description of Modifying Procedure

Files were downloaded and extracted to the working directory. Subject ID (from subject_train.txt/subject_test.txt), feature data (from X_train.txt/X_test.txt), and activity labels (from y_train.txt/y_test.txt) for each of training and test data were obtained from given files and column-bound together to form separate training and test sets. These separate training and test sets were then row-bound together to produce a single data frame with columns being features and rows being id-measurement period records.

Column labels were extracted from features.txt and added as column labels. The large dataset was then filtered and only those mean and standard deviation features that were most proximal to the original measurements of tri-axial acceleration and angular velocity were retained (temporal measurements of body/gravity acceleration and body angular velocity in x, y, and z directions). Target activity labels (integers) were then transformed to descriptive activity name (strings) factor variables. Finally, the resulting data set was melted along the variables "id" and "activity" so that a final dataset of mean values for the feature vectors across "id" and "activity" could be produced.

Variables

id
: subject identity; integer: 1-30

tBodyAcc-mean()-X
: temporal body acceleration means in x-coordinate direction

tBodyAcc-mean()-Y
: temporal body acceleration means in y-coordinate direction

tBodyAcc-mean()-Z
: temporal body acceleration means in z-coordinate direction

tBodyAcc-std()-X
: temporal body acceleration standard deviations in x-coordinate direction

tBodyAcc-std()-Y
: temporal body acceleration standard deviations in y-coordinate direction

tBodyAcc-std()-Z
: temporal body acceleration standard deviations in z-coordinate direction

tGravityAcc-mean()-X
: temporal gravity acceleration means in x-coordinate direction

tGravityAcc-mean()-Y
temporal gravity acceleration means in y-coordinate direction

tGravityAcc-mean()-Z
temporal gravity acceleration means in z-coordinate direction

tGravityAcc-std()-X
temporal gravity acceleration standard deviation in x-coordinate direction

tGravityAcc-std()-Y
temporal gravity acceleration standard deviation in y-coordinate direction

tGravityAcc-std()-Z
temporal gravity acceleration standard deviation in z-coordinate direction

tBodyGyro-mean()-X
temporal angular velocity means in x-coordinate direction

tBodyGyro-mean()-Y
temporal angular velocity means in y-coordinate direction

tBodyGyro-mean()-Z
temporal angular velocity means in z-coordinate direction

tBodyGyro-std()-X
temporal angular velocity standard deviations in x-coordinate direction

tBodyGyro-std()-Y
temporal angular velocity standard deviations in y-coordinate direction

tBodyGyro-std()-Z
temporal angular velocity standard deviations in z-coordinate direction

activity
WALKING
WALKING_UPSTAIRS
WALKING_DOWNSTAIRS
SITTING
STANDING
LAYING