Codebook for Tidydata.txt

The Data in this dataset is derived from data produced by Reyes-Ortiz, Ghio, Oneto, and Anguita and presented at the International Workshop of Ambient Assisted Living (IWAAL 2012). Vitoria-Gasteiz, Spain. Dec 2012

The original dataset collected data from 30 subjects performing 6 activities whilst wearing a Samsung Galaxy SII Smartphone. Data recorded from sensors in the smartphone was used to create a 561 feature vector for each observation.

The Tidydata.txt file extracts information on a subset of 66 of these 561 features and these are summarised by subject and activity to give a dataset with 180 observations.

Original Authors description of study:

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.

Description of Tidydata dataset.

Classification Variables:

There are two variables which are used to classify the data. These are the Activity and Subject variables. There is one record for each combination of Activity and Data.

Activity

There are 6 activities as follows: There is one record for each activity carried out by each participant.

WALKING, WALKING_UPSTAIRS, WALKING_DOWNSTAIRS, SITTING, STANDING, LAYING

Subject There were 30 subjects. Each subject was assigned a number from 1 to 30.

There are then a number of sensors in the Samsung SII phones which were recorded:

-Measurements are either recorded in time domain or frequency domain.

Analysis Variables:

For the purpose of this project 33 analysis variables were taken from the original. These numbers were then summarised to create a single record for each of the 6 activities by each of the 30 subjects with means of the (33 means and 33 standard deviations) individual analysis variables from the original dataset.

The names of the variables are similar to but not identical to those in the original dataset as some cleansing and standardisation of naming was carried out on the original variable names.

The analysis variables which were chosen from the original dataset were those which had both means and standard deviations.

These Variables are assembled from a number of categories:

Time Domain/ Frequency Domain Indicated by the prefix t or f.

Body = estimated Body effects

Gravity= Estimated Gravitational Effects

Gyro = Gyroscope

Acc=Acceleration (Acceleration is rate of change of velocity)

Jerk= Jerk (Jerk is rate of change of acceleration)

Axis= X,Y,Z where the components of a measurement could be spilt into X, Y and Z axis then a separate measurement is given for each axis.

Statistic= Mean or STD (Standard Deviation).

A summary table below shows how each variable maps to the various parameters:

Variable Code	Time/ Frequency Domain	Body	Gravity	Gyro	Accelerometer	Jerk	Mag	Axis	Statistic
tBodyAcc-X_Mean	t	✓			✓			X	Mean
tBodyAcc-Y_Mean	t	✓			✓			Y	Mean
tBodyAcc-Z_Mean	t	✓			✓			Z	Mean
tBodyAcc-X_STD	t	✓			✓			X	STD
tBodyAcc-Y_STD	t	✓			✓			Y	STD
tBodyAcc-Z_STD	t	✓			✓			Z	STD
tGravityAcc-X_Mean	t		✓		✓			X	Mean
tGravityAcc-Y_Mean	t		✓		✓			Y	Mean
tGravityAcc-Z_Mean	t		✓		✓			Z	Mean
tGravityAcc-X_STD	t		✓		✓			X	STD
tGravityAcc-Y_STD	t		✓		✓			Y	STD
tGravityAcc-Z_STD	t		✓		✓			Z	STD
tBodyAccJerk-X_Mean	t	✓			✓	✓		X	Mean
tBodyAccJerk-Y_Mean	t	✓			✓	✓		Y	Mean
tBodyAccJerk-Z_Mean	t	✓			✓	✓		Z	Mean
tBodyAccJerk-X_STD	t	✓			✓	✓		X	STD
tBodyAccJerk-Y_STD	t	✓			✓	✓		Y	STD
tBodyAccJerk-Z_STD	t	✓			✓	✓		Z	STD
tBodyGyro-X_Mean	t	✓		✓				X	Mean
tBodyGyro-Y_Mean	t	✓		✓				Y	Mean
tBodyGyro-Z_Mean	t	✓		✓				Z	Mean
tBodyGyro-X_STD	t	✓		✓				X	STD
tBodyGyro-Y_STD	t	✓		✓				Y	STD
tBodyGyro-Z_STD	t	✓		✓				Z	STD
tBodyGyroJerk-X_Mean	t	✓		✓		✓		X	Mean
tBodyGyroJerk-Y_Mean	t	✓		✓		✓		Y	Mean
tBodyGyroJerk-Z_Mean	t	✓		✓		✓		Z	Mean
tBodyGyroJerk-X_STD	t	✓		✓		✓		X	STD
tBodyGyroJerk-Y_STD	t	✓		✓		✓		Y	STD
tBodyGyroJerk-Z_STD	t	✓		✓		✓		Z	STD
tBodyAccMag_Mean	t	✓			✓		✓		Mean
tBodyAccMag_STD	t	✓			✓		✓		STD

Variable Code	Time/ Frequency Domain	Body	Gravity	Gyro	Accelerometer	Jerk	Mag	Axis	Statistic
tGravityAccMag_Mean	t		✓		✓		✓		Mean
tGravityAccMag_STD	t		✓		✓		✓		STD
tBodyAccJerkMag_Mean	t	✓			✓	✓	✓		Mean
tBodyAccJerkMag_STD	t	✓			✓	✓	✓		STD
tBodyGyroMag_Mean	t	✓		✓			✓		Mean
tBodyGyroMag_STD	t	✓		✓			✓		STD
tBodyGyroJerkMag_Mean	t	✓		✓		✓	✓		Mean
tBodyGyroJerkMag_STD	t	✓		✓		✓	✓		STD
fBodyAcc-X_Mean	f	✓			✓			X	Mean
fBodyAcc-Y_Mean	f	✓			✓			Y	Mean
fBodyAcc-Z_Mean	f	✓			✓			Z	Mean
fBodyAcc-X_STD	f	✓			✓			X	STD
fBodyAcc-Y_STD	f	✓			✓			Y	STD
fBodyAcc-Z_STD	f	✓			✓			Z	STD
fBodyAccJerk-X_Mean	f	✓			✓	✓		X	Mean
fBodyAccJerk-Y_Mean	f	✓			✓	✓		Y	Mean
fBodyAccJerk-Z_Mean	f	✓			✓	✓		Z	Mean
fBodyAccJerk-X_STD	f	✓			✓	✓		X	STD
fBodyAccJerk-Y_STD	f	✓			✓	✓		Y	STD
fBodyAccJerk-Z_STD	f	✓			✓	✓		Z	STD
fBodyGyro-X_Mean	f	✓		✓				X	Mean
fBodyGyro-Y_Mean	f	✓		✓				Y	Mean
fBodyGyro-Z_Mean	f	✓		✓				Z	Mean
fBodyGyro-X_STD	f	✓		✓				X	STD
fBodyGyro-Y_STD	f	✓		✓				Y	STD
fBodyGyro-Z_STD	f	✓		✓				Z	STD
fBodyAccMag_Mean	f	✓			✓		✓		Mean
fBodyAccMag_STD	f	✓			✓		✓		STD
fBodyAccJerkMag_Mean	f	✓			✓	✓	✓		Mean
fBodyAccJerkMag_STD	f	✓			✓	✓	✓		STD
fBodyGyroMag_Mean	f	✓		✓			✓		Mean
fBodyGyroMag_STD	f	✓		✓			✓		STD
fBodyGyroJerkMag_Mean	f	✓		✓		✓	✓		Mean
fBodyGyroJerkMag_STD	f	✓		✓		✓	✓		STD

Information on original datasets:

License:

========

[1] Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra and Jorge L. Reyes-Ortiz. Human Activity Recognition on Smartphones using a Multiclass Hardware-Friendly Support Vector Machine. International Workshop of Ambient Assisted Living (IWAAL 2012). Vitoria-Gasteiz, Spain. Dec 2012

Jorge L. Reyes-Ortiz, Alessandro Ghio, Luca Oneto, Davide Anguita. November 2012.