Codebook: Getting and Cleaning Data Course Project

Data Set Codebook:

Prepared by Paul H

A full description of the experiment and data collected can be found at:

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

Summary

Data was collected on 30 volunteers wearing a Samsung Galaxy S and data collected from the embedded accelerometer and gyroscope. Data was collected on six activities: WALKING, WALKING\_UPSTAIRS, WALKING\_DOWNSTAIRS, SITTING, STANDING, LAYING. Seventy percent of the volunteers were used for generating training data and 30% for testing. Further information can be found at above web site. 561 attributes were generated for each activity giving 10299 data points.

The data was provided as a zip file available from:

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

Files included with the zip include:

README.txt

Features\_info.txt – brief description of the features vector variables

The following files were used to build the dataset.

Activity\_labels.txt – the six activities and numeric key(1-6)

Features.txt – list of 561 attributes and numeric key(1-561)

Subject\_train.txt – Human subjects numeric key corresponding to rows in X\_train.txt

train/X\_train.txt -Training set of 561 attributes

train/y\_train.txt: Training activities numeric key corresponding to rows in X\_train.txt

Subject\_test.txt – Human subjects numeric key corresponding to rows in X\_test.txt

test/X\_test.txt: Test set of 561 attributes .

test/y\_test.tx': Test activities numeric key corresponding to rows in X\_train.txt

Nine other files each for training and testing were available, but not required for the this project.

Tidy Data Set variables

The tidy data set is 180 rows x 68 columns. Cols 1 and 2 are the Subject and Activity, respectively. The remaining 66 columns are the mean and standard deviation columns selected from the original 561 columns. The rows correspond to the 30 subjects x 6 activities with the data for each attribute averaged over the respective activity.

APPENDIX

Readme.txt

==================================================================

Human Activity Recognition Using Smartphones Dataset

Version 1.0

==================================================================

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

Smartlab - Non Linear Complex Systems Laboratory

DITEN - Università degli Studi di Genova.

Via Opera Pia 11A, I-16145, Genoa, Italy.

activityrecognition@smartlab.ws

www.smartlab.ws

==================================================================

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.

Notes:

- Features are normalized and bounded within [-1,1].

- Each feature vector is a row on the text file.

For more information about this dataset contact: activityrecognition@smartlab.ws

License:

Use of this dataset in publications must be acknowledged by referencing the following publication [1]

[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

This dataset is distributed AS-IS and no responsibility implied or explicit can be addressed to the authors or their institutions for its use or misuse. Any commercial use is prohibited.

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