**Codebook**

**Human Activity Recognition Using Smartphones Data Set**

This codebook describes the variables, the data, and any transformations or work that are performed to clean the data.

**The data source**

* Original data is available at:

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

* Original description of the dataset is available at:

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

**Data Set Information**

The Human Activity Recognition Using Smartphones 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.

**The data**

The original 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.

**Aims of the transformations**

There are 5 parts:

1. Merges the training and the test sets to create one data set.
2. Extracts only the measurements on the mean and standard deviation for each measurement.
3. Uses descriptive activity names to name the activities in the data set
4. Appropriately labels the data set with descriptive activity names.
5. Creates a second, independent tidy data set with the average of each variable for each activity and each subject.

**How “run\_analysis.R” implements the transformation aims:**

After data source downloaded into working directory:

* Required packages “reshapre2” and “data.table”, “dplyr” loaded.
* feutures.txt file is loaded, and the positions and labels of required data columns taken from it.
* Activity labels are loaded from the activity\_labels.txt file.
* Activity Id for test group data loaded from /test/y\_test.txt, and merged with activity labels.
* Test group data loaded from /test/X\_test.txt .
* Only required columns of test group data is exteracted and kept, and each feature name assigned to the selected columns.
* Subject IDs for test group data loaded from test/subject\_test.txt.
* A new data table created by merging (column-wise) the subject ID, activity labels, and test group datasets.
* Columns are renamed (labelled) appropriately for later processes and with descriptive variable names.
* All the above steps are also performed on the train group data. Note that the train data loaded from respected file and folders responding to the train datasets.
* Both test and train datasets are merged together (raw-wise) to create the merged dataset.
* To create tidy data from the merged data, the merged data first melted to a tall data table which represents all the combination of each subject, each activity and each feature variable.
* Finally, the melted data is summarised to show the average of each variable for each activity and each subject.