Code Book for Human Activity Recognition Using Smartphones Data Set

The Ohio State University
April 8, 2020

Pre-processing steps to download and read data
1. Data was downloaded and unzipped as UCI HAR Dataset to working directory.
2. Six variables were created to read each the separate tables within the train and test folders:
    x_train, y_train, subject_train, x_test, y_test, subject_test
3. Two additional variables were created from features and activity_labels files:
    features, activities

The run_analysis.R script performs data processing and manipulation as outlined in the 5 learning objectives for the course project.

1. Merges the training and the test sets to create one data set.
    x_comb is created by merging x_train to x_test with rbind() function
    y_comb is created by merging y_train to y_test with rbind()
    subject_comb is created by merging subject_train to subject_test with rbind()
    all_data is created by merging subject_comb, x_comb, y_comb with cbind()

2. Extracts only the measurements on the mean and standard deviation for each measurement.
    neat_data is created by selecting the subject, shorthand columns and those that contain the mean and std (standard deviation) measurements

3. Uses descriptive activity names to name the activities in the data set.
    All numbers in the “shorthand” column of neat_data was replaced with the corresponding activity from the second column of the activities variable

4. Appropriately labels the data set with descriptive variable names
    column 1 of neat_data was renamed to “Subject”
    column 2 was renamed to “Activity”
    All “Acc” instances were replaced with “Accelerometer”
    All “Gyro” were replaced with “Gyroscope”
    All “Mag” were replaced with “Magnitude”
    All “tGravity” was replaced with “TimeGravity”
    All “freq” was replaced with “Frequency”
    All “tBody” was replaced with “TimeBody”
    All “f” was replaced with “Frequency”

5. From the data set in step 4, creates a second, independent tidy data set with the average of each variable for each activity and each subject

Final_neat_data is created by grouping the data by “Subject” and “Activity” and then summarizing all variables with the mean
Data is exported as a text file to My_Neat_Data.txt

Data for this analysis was sourced from:
Jorge L. Reyes-Ortiz(1,2), Davide Anguita(1), Alessandro Ghio(1), Luca Oneto(1) and Xavier Parra(2) 1 - Smartlab - Non-Linear Complex Systems Laboratory DITEN - Università degli Studi di Genova, Genoa (I-16145), Italy. 2 - CETpD - Technical Research Centre for Dependency Care and Autonomous Living Universitat Politècnica de Catalunya (BarcelonaTech). Vilanova i la Geltrú (08800), Spain activityrecognition '@' smartlab.ws
Accessed: April 7, 2020

More information regarding this data can be found at
http://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones