CodeBook

Guohong He

11/11/2020

Code Book

This code book will describe the processing used in this course project to create a tidy data.

Overview

30 volunteers performed 6 different activities while wearing a smartphone in their waist. The data of sensors (accelerometers and gyroscopes embedded in smartphone) are collected for human movement recognitions.

Raw data files used

features.txt: Names of the 561 variables

activity_labels.txt: Names and IDs for each of the 6 activities

X_test.txt: 2947 observations of the 561 features, for 9 of the 30 volunteers

y_test.txt: A vector of 2947 integers, denoting the ID of the activity of the observations

subject_test.txt: A vector of 2947 integers, denoting the ID of the volunteer of the observations

X_train.txt: 7352 observations of the 561 features, for 21 of the 30 volunteers

y_train.txt: A vector of 7352 integers, denoting the ID of the activity of the observations

subject_train.txt: A vector of 7352 integers, denoting the ID of the volunteer of the observations

Processing steps

1. Input all relevant data files. Add appropriate column headers. Merge train set and test set as ***"X_all"***
2. Select those the measurements with key words ***"mean"*** or ***"std"***. Get data set ***"tidyData"***
3. Add two addition columns ***"subjects"*** and ***"activities"***. The content of "activities" column had been replaced by meaningful words.
4. The feature names provided in data sets are not clear. Change first letter ***"t"*** as ***"time"***, ***"f"*** as ***"freq"***, and move ***"mean"*** and ***"std"*** to the end of variable names. A name reference file ***"varNames.txt"*** is exported.
5. Create an independent tidy data set with the average of each variable for each activity and each subject. The data set is ***"tidyData"*** and it is exported as ***"tidyData.txt"***.

The data processing R script ***"run_analysis.R"***, tidy data set ***"tidyData.txt"*** and feature reference list ***"varNames.txt"*** are put in this repo.