Human Activity Recognition Using Smartphones Tidy Dataset

Daniel Fynes-Clinton

Introduction

This readme file details the tidy dataset derived from the *Human Activity Recognition Using Smartphones* dataset, as well as the R code which was used to extract the tidy dataset (**run_analysis.R**).

Tidy Dataset

The tidy dataset (**tidyData.txt**) contains 68 columns and 180 rows of data. The columns consist of 66 measurement/dependent variables are average values, grouped by the two independent variables (Activity and Subject Identifier Number). The measurement variables consist of the mean and standard deviation of a variety of acceleration and angular velocity measurements in three dimensions (X,Y,Z), in the time and frequency domains. Each variable is described further in the codebook (**Codebook.md**).

R code (run_analysis.R)

The **run_analysis.R** code performs the following functions:

1. Extracts the label data for the different activities and features (measurements) from their respective text files.
2. Extracts and combines (using the column bind function) the subject, activity and measurement data from the train and test datasets from their respective text files.
3. The train and test datasets are then merged together using the row bind function.
4. Only the mean and standard deviation measurements are extracted using regular expression matching and the *grep* function.
5. Descriptive activity names from step 1 are merged into the dataset, and used in place of the activity indices (1...6).
6. The measurement variable names are modified to be more descriptive/readable. A number of substitute functions are performed to achieve this, for example subsituting *Time domain signal* instead of t.
7. From the dataset in step 5, a second, independent tidy dataset is extracted, with the average of each variable grouped by activity and subject. This is achieved using the *group_by* and *summarize* functions of the **dplyr** library.
8. The final dataset is written to a file (**tidyData.txt**) using *write.table*.

The dataset includes the following files:

-README.md

-run_analysis.R

-Codebook.pdf

-tidyData.txt