Codebook for UCIHARtidy1.txt

Variable descriptors

For brevity, these are described in groups where similar. The descriptions are based on the document `features_info.txt`.

`activity`

Factor, 6 levels. Identifies activity being undertaken. Walk/Walkup/Walkdown/Stand/Sit/Lay

`subject`

Integer 1 to 30. Identifies test subject.

`tbodyaccmeanx`
`tbodyaccmeany`
`tbodyaccmeanz`
`tbodyaccstdx`
`tbodyaccstdy`
`tbodyaccstd`

mean and std deviations, body acceleration x, y, z axes

`tgravityaccmeanx`
`tgravityaccmeany`
`tgravityaccmeanz`
`tgravityaccstdx`
`tgravityaccstdy`
`tgravityaccstdz`

mean and std deviations, gravity acceleration x, y, z axes

`tbodyaccjerkmeanx`
`tbodyaccjerkmeany`
`tbodyaccjerkmeanz`
`tbodyaccjerkstdx`
`tbodyaccjerkstdy`
`tbodyaccjerkstdz`

mean and std deviations, body linear jerk x, y, z axes

`tbodygyromeanx`
`tbodygyromeany`
`tbodygyromeanz`
`tbodygyrostdx`
`tbodygyrostdy`
`tbodygyrostdz`

mean and std deviations, body gyroscope raw signals x, y, z axes

`tbodygyrojerkmeanx`
`tbodygyrojerkmeany`
`tbodygyrojerkmeanz`
`tbodygyrojerkstdx`
`tbodygyrojerkstdy`
`tbodygyrojerkstdz`

mean and std deviations, body angular x, y, z axes

tbodyaccmagmean
tbodyaccmagstd
tgravityaccmagmean
tgravityaccmagstd
tbodyaccjerkmagmean
tbodyaccjerkmagstd
tbodygyromagmean
tbodygyromagstd
tbodygyrojerkmagmean
tbodygyrojerkmagstd

mean and std deviation, magnitudes

fbodyaccmeanx
fbodyaccmeany
fbodyaccmeanz
fbodyaccstdx
fbodyaccstdy
fbodyaccstdz
fbodyaccmeanfreqx
fbodyaccmeanfreqy
fbodyaccmeanfreqz
fbodyaccjerkmeanx
fbodyaccjerkmeany
fbodyaccjerkmeanz
fbodyaccjerkstdx
fbodyaccjerkstdy
fbodyaccjerkstdz
fbodyaccjerkmeanfreqx
fbodyaccjerkmeanfreqy
fbodyaccjerkmeanfreqz
fbodygyromeanx
fbodygyromeany
fbodygyromeanz
fbodygyrostdx
fbodygyrostdy
fbodygyrostdz
fbodygyromeanfreqx
fbodygyromeanfreqy
fbodygyromeanfreqz
fbodyaccmagmean
fbodyaccmagstd
fbodyaccmagmeanfreq
fbodybodyaccjerkmagmean
fbodybodyaccjerkmagstd
fbodybodyaccjerkmagmeanfreq
fbodybodygyromagmean
fbodybodygyromagstd
fbodybodygyromagmeanfreq
fbodybodygyrojerkmagmean
fbodybodygyrojerkmagstd
fbodybodygyrojerkmagmeanfreq

mean and std deviations, fast fourier transforms

`angletbodyaccmeangravity`
`angletbodyaccjerkmeangravitymean`
`angletbodygyromeangravitymean`
`angletbodygyrojerkmeangravitymean`
`anglexgravitymean`
`angleygravitymean`
`anglezgravitymean`
mean and std deviation angle between vectors

Creation of the tidy dataset

The dataset was created from the documents in UCI HAR Dataset by applying the R script appended below. It works as follows:

The file run_analysis.R requires that the uncompressed folder "UCI HAR Dataset" be in the working directory for R, wherever that may be. If in doubt run getwd() and place the "UCI HAR Dataset" folder in the directory identified. HOW DO YOU RUN IT??

The R script will then:

- read the file X_train.txt - this file contains the actual data, 561 variables, 7352 observations
- read the file features.txt. The second column is used to name the variables in X-train.txt
- a search is run in the features labels for all variables containing 'mean' or 'std', and this is used to subset the X_train dataset. This reduces it to 86 columns.
- read the file y_train.txt - this file describes the activity being undertaken, 7352 rows. Name this 'activity' and add as a column to X_train.
- read the file subject_train.txt - this identifies the training subject (1 to 30). 7352 rows. Name this 'subject', and add as another column.
- Create a dataset with descriptive activity labels and merges this with the main dataset. This dataframe (trainmean)now has 88 columns (data plus activity plus subject).

The script will then perform a similar series of steps on the files in the 'test' folder, to create a dataframe called merged2.

The script will ignore the data in the Inertial Signals folder. Since these data do not contain any information on means or standard deviations, none of these figures are needed in the final tidy dataset. Not including them cuts out unnecessary steps and makes the scripts run faster.

merged1 and merged2 are combined with rbind() to create merged3, which has 88 columns and 7352 + 2947 = 10299 rows. The variable names are tidied up to remove caps, commas, and spurious brackets. This is saved as `UCIHARtidy1.txt`.

This is then grouped and summarized (means) and saved as `UCIHARtidy2.txt`