Data Cleaning Project Read Me file

Doug Joubert

October 9, 2019

Table of Contents

This DATACLEANINGPROJECT readme.txt file was generated on 2019-10-09 by Douglas Joubert

# General Information

* This repository was created as part of the Getting and Cleaning Data course project. It has the instructions on how to run analysis on Human Activity recognition dataset.

## Author Information

Creator: Douglas Joubert

## Title of Dataset:

### Data set information

The experiments were carried out with a group of 30 volunteers within an age bracket of 19-48 years. They performed a protocol of activities composed of six basic activities: three static postures (standing, sitting, lying) and three dynamic activities (walking, walking downstairs and walking upstairs). The experiment also included postural transitions that occurred between the static postures. These are: stand-to-sit, sit-to-stand, sit-to-lie, lie-to-sit, stand-to-lie, and lie-to-stand. All the participants were wearing a smartphone (Samsung Galaxy S II) on the waist during the experiment execution. We captured 3-axial linear acceleration and 3-axial angular velocity at a constant rate of 50Hz using the embedded accelerometer and gyroscope of the device. The experiments were video-recorded to label the data manually. The obtained dataset was randomly partitioned into two sets, where 70% of the volunteers was selected for generating the training data and 30% the test data [@Anguita].

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 (Anguita, 2013).

Check the README.md for further details about this dataset.

### Date of data collection

2012-12-10, more information available from Data Set Description [file](http://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones#).

### Attribute Information:

For each record in the dataset it is provided: \* Triaxial acceleration from the accelerometer (total acceleration) and the estimated body acceleration. \* Triaxial Angular velocity from the gyroscope. \* A 561-feature vector with time and frequency domain variables. \* Its activity label. \* An identifier of the subject who carried out the experiment.

# Code style and Data used

* This project was written in *R version 3.5.1 (2018-07-02))*
* Full description of the data used and analysis performed is found in the CodeBook.md or Codebook.rmd files.

# References

1. Anguita, D., Ghio, A., Oneto, L., Parra, X., & Reyes-Ortiz, J. L. (2013). Human Activity Recognition Using Smartphones Data Set. Retrieved October 4, 2019, from <http://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones#>

## SHARING/ACCESS INFORMATION

Licenses/restrictions placed on the data, or limitations of reuse:

Recommended citation for the data:

Citation for and links to publications that cite or use the data:

Links to other publicly accessible locations of the data:

Links/relationships to ancillary or related data sets:

|  |
| --- |
| DATA & FILE OVERVIEW |

File list (filenames, directory structure (for zipped files) and brief description of all data files):

Relationship between files, if important for context:

Additional related data collected that was not included in the current data package:

If data was derived from another source, list source:

If there are there multiple versions of the dataset, list the file updated, when and why update was made:

|  |
| --- |
| METHODOLOGICAL INFORMATION |

Description of methods used for collection/generation of data:

Methods for processing the data:

Software- or Instrument-specific information needed to interpret the data, including software and hardware version numbers:

Standards and calibration information, if appropriate:

Environmental/experimental conditions:

Describe any quality-assurance procedures performed on the data:

People involved with sample collection, processing, analysis and/or submission:

|  |
| --- |
| DATA-SPECIFIC INFORMATION |

Number of variables:

Number of cases/rows:

Variable list, defining any abbreviations, units of measure, codes or symbols used:

Missing data codes:

Specialized formats or other abbreviations used: