Codebook

Przemysław Skowron

8/6/2020

This codebook describes list of all variables, data and any transformations that were applied to source data to produce tidy dataset stored in output_data.txt file.

Experiment and source data description

The experiments have been carried out with a group of 30 volunteers within an age bracket of 19-48 years. Each person performed six activities wearing a smartphone (Samsung Galaxy S II) on the waist. Using its embedded accelerometer and gyroscope, we captured 3-axial linear acceleration and 3-axial angular velocity at a constant rate of 50Hz. The experiments have been video-recorded to label the data manually. The obtained dataset has been randomly partitioned into two sets, where 70% of the volunteers was selected for generating the training data and 30% the test data.

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.

For each record in the source 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.

Transformation

The following steps were applied to the source dataset to transform it into tidy output dataset (please refer fo README.Rmd file for more details on the source code and algorythm used):

1. Training and Test datasets were merged into one dataset.
2. Out of 561 features only measurements on mean and standard deviations were extracted (66 variables).
3. Activity id's were replaced with descriptive activity names and added to the dataset.
4. Each variable was provided with descriptive label following the same naming convention (e.g. timeGravityAccMeanX)
5. A summary dataset was constructed with the average of each variable for each subject and each activity.
6. The summary dataset has been saved as space-delimited flat text file: output_data.txt

Output dataset

Description

The output dataset is a data frame with 180 rows and 68 columns. There are no missing values (NA's) in the dataset. The dataset is tidy as per tidy dataset definition.

Each row represents the vector of averagees of each variable for a given subject and activity. As there are 30 subjects and 6 activities the number of rows is 30 * 6 = 180.

The first 2 columns in the dataset are "identifiers":

1. **subjectId** representing the subject (person) in the experiment. It is integer variable with values in the range between 1 and 30.
2. **activityName** is a desctiptive name of activity performed by each subject during the experiment. It is factor variable with the following levels:

Table 1: Factor levels for activityName variable:

activityName value	Value description
LAYING	Subject was laying down during the experiment.
SITTING	Subject was sitting during the experiment.
STANDING	Subject was standing during the experiment.
WALKING	Subject was walking on flat surface during the experiment.
WALKING_DOWNSTAIRS	Subject was walking downstairs during the experiment.
WALKING_UPSTAIRS	Subject was walking upstairs during the experiment.

The following 66 columns in the dataset represent average measurements of 66 variables per activity per subject. They are all numeric variables. Each variable has descriptive name foillowing the same naming convention consisting of multiple elements, e.g:

- **time** or **freq** prefix indicates time or frequency domain signals;
- **Acc** or **Gyro** strings indicate that signal comes either from accelerometer or gyroscope;
- **Body** or **Gravity** strings indicate body and gravity acceleration signals;
- **Jerk** string indicates calculated Jerk signals;
- **Mag** string indicates magnitude of three-dimensional signals calculated using the Euclidean norm;
- **Mean** or **Std** strings indicated mean or standard deviation variable;
- **X**, **Y** or **Z** suffix is used to denote 3-axial signals in the X, Y and Z directions.

List of variables

The following list shows names of all 68 variables (indicators: SubjectId, activityName and 66 numeric variables) along with some statistics for each variable (class, number of unique values, rounded min and max value) in the order as they appear in the dataset:

```
##                      Variable     Class Unique_values Min_value        Max_value
## 1                   subjectId   integer            30         1               30
## 2                activityName character             6    LAYING WALKING_UPSTAIRS
## 3             timeBodyAccMeanX   numeric           180    0.2216           0.3015
## 4             timeBodyAccMeanY   numeric           180   -0.0405          -0.0013
## 5             timeBodyAccMeanZ   numeric           180   -0.1525          -0.0754
## 6              timeBodyAccStdX   numeric           180   -0.9961           0.6269
## 7              timeBodyAccStdY   numeric           180   -0.9902           0.6169
## 8              timeBodyAccStdZ   numeric           180   -0.9877            0.609
## 9          timeGravityAccMeanX   numeric           180     -0.68           0.9745
## 10         timeGravityAccMeanY   numeric           180   -0.4799           0.9566
```

```
## 11          timeGravityAccMeanZ   numeric           180   -0.4951           0.9579
## 12           timeGravityAccStdX   numeric           180   -0.9968          -0.8296
## 13           timeGravityAccStdY   numeric           180   -0.9942          -0.6436
## 14           timeGravityAccStdZ   numeric           180    -0.991          -0.6102
## 15         timeBodyAccJerkMeanX   numeric           180    0.0427           0.1302
## 16         timeBodyAccJerkMeanY   numeric           180   -0.0387           0.0568
## 17         timeBodyAccJerkMeanZ   numeric           180   -0.0675           0.0381
## 18          timeBodyAccJerkStdX   numeric           180   -0.9946           0.5443
## 19          timeBodyAccJerkStdY   numeric           180   -0.9895           0.3553
## 20          timeBodyAccJerkStdZ   numeric           180   -0.9933            0.031
## 21            timeBodyGyroMeanX   numeric           180   -0.2058           0.1927
## 22            timeBodyGyroMeanY   numeric           180   -0.2042           0.0275
## 23            timeBodyGyroMeanZ   numeric           180   -0.0725           0.1791
## 24             timeBodyGyroStdX   numeric           180   -0.9943           0.2677
## 25             timeBodyGyroStdY   numeric           180   -0.9942           0.4765
## 26             timeBodyGyroStdZ   numeric           180   -0.9855           0.5649
## 27       timeBodyGyroJerkMeanX   numeric           180   -0.1572          -0.0221
## 28       timeBodyGyroJerkMeanY   numeric           180   -0.0768          -0.0132
## 29       timeBodyGyroJerkMeanZ   numeric           180   -0.0925          -0.0069
## 30        timeBodyGyroJerkStdX   numeric           180   -0.9965           0.1791
## 31        timeBodyGyroJerkStdY   numeric           180   -0.9971           0.2959
## 32        timeBodyGyroJerkStdZ   numeric           180   -0.9954           0.1932
## 33          timeBodyAccMagMean   numeric           180   -0.9865           0.6446
## 34           timeBodyAccMagStd   numeric           180   -0.9865           0.4284
## 35       timeGravityAccMagMean   numeric           180   -0.9865           0.6446
## 36        timeGravityAccMagStd   numeric           180   -0.9865           0.4284
## 37      timeBodyAccJerkMagMean   numeric           180   -0.9928           0.4345
## 38       timeBodyAccJerkMagStd   numeric           180   -0.9946           0.4506
## 39         timeBodyGyroMagMean   numeric           180   -0.9807            0.418
## 40          timeBodyGyroMagStd   numeric           180   -0.9814              0.3
## 41     timeBodyGyroJerkMagMean   numeric           180   -0.9973           0.0876
## 42      timeBodyGyroJerkMagStd   numeric           180   -0.9977           0.2502
## 43            freqBodyAccMeanX   numeric           180   -0.9952            0.537
## 44            freqBodyAccMeanY   numeric           180    -0.989           0.5242
## 45            freqBodyAccMeanZ   numeric           180   -0.9895           0.2807
## 46             freqBodyAccStdX   numeric           180   -0.9966           0.6585
## 47             freqBodyAccStdY   numeric           180   -0.9907           0.5602
## 48             freqBodyAccStdZ   numeric           180   -0.9872           0.6871
## 49       freqBodyAccJerkMeanX   numeric           180   -0.9946           0.4743
## 50       freqBodyAccJerkMeanY   numeric           180   -0.9894           0.2767
## 51       freqBodyAccJerkMeanZ   numeric           180    -0.992           0.1578
## 52        freqBodyAccJerkStdX   numeric           180   -0.9951           0.4768
## 53        freqBodyAccJerkStdY   numeric           180   -0.9905           0.3498
## 54        freqBodyAccJerkStdZ   numeric           180   -0.9931          -0.0062
## 55          freqBodyGyroMeanX   numeric           180   -0.9931            0.475
## 56          freqBodyGyroMeanY   numeric           180    -0.994           0.3288
## 57          freqBodyGyroMeanZ   numeric           180    -0.986           0.4924
## 58           freqBodyGyroStdX   numeric           180   -0.9947           0.1966
## 59           freqBodyGyroStdY   numeric           180   -0.9944           0.6462
## 60           freqBodyGyroStdZ   numeric           180   -0.9867           0.5225
## 61         freqBodyAccMagMean   numeric           180   -0.9868           0.5866
## 62          freqBodyAccMagStd   numeric           180   -0.9876           0.1787
## 63 freqBodyBodyAccJerkMagMean   numeric           180    -0.994           0.5384
## 64  freqBodyBodyAccJerkMagStd   numeric           180   -0.9944           0.3163
```

```
## 65     freqBodyBodyGyroMagMean   numeric           180   -0.9865           0.204
## 66      freqBodyBodyGyroMagStd   numeric           180   -0.9815          0.2367
## 67 freqBodyBodyGyroJerkMagMean   numeric           180   -0.9976          0.1466
## 68  freqBodyBodyGyroJerkMagStd   numeric           180   -0.9976          0.2878
```