CODEBOOK

UNIVERSITY OF CALIFORNIA IRVINE MACHINE LEARNING REPOSITORY

SMARTPHONES HUMAN ACTIVITY RECOGNITION DATA SET

June 2014

Getting and Cleaning Data

Coursera: Johns Hopkins Bloomberg School of Public Health

DATA COLLECTION AND SCRIPT DESCRIPTION

SUMMARY: The purpose of this data collection and script is to construct a composite data set from the University of California Machine Learning Repository (UCI MLR, 2012) dataset, consisting of training and test group measurement statistics for six specific activities.

SOURCE DATA AND STUDY DESIGN:

Source data were obtained from the UCI MLR "Human Activity Recognition Using Smartphones Data Set" (UCI MLR, 2012)(http://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones).

The source data zip file is at https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset.zip.

The Human Activity Recognition database is built from the recordings of 30 subjects performing activities of daily living while carrying a waist-mounted smartphone with embedded inertial sensors.

As described by UCI MRL (2012):

"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 (WALKING, WALKING\_UPSTAIRS, WALKING\_DOWNSTAIRS, SITTING, STANDING, LAYING) 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."

A composite data set was constructed from the UCI MCL (2012) dataset, consisting of training and test group measurement statistics for six specific activities. The composite data set was constructed from the following data files included in the UCI MCL (2012) database:

| File | Variables | Records | Description | Values |
| --- | --- | --- | --- | --- |
| activity\_labels.txt | 2 | 6 | Links activity index with activity name | WALKING  WALKING\_UPSTAIRS  WALKING\_DOWNSTAIRS  SITTING  STANDING  LAYING |
| features.txt | 2 | 561 | Index and List of all features: Statistical variables for each feature vector (BodyAcc, GravityAcc, BodyGyro), in the x, y, and z directions, for time (t) and frequency (f) domain signals, and angle variables. | See features\_info.txt  Feature vectors:  tBodyAcc-XYZ  tGravityAcc-XYZ  tBodyAccJerk-XYZ  tBodyGyro-XYZ  tBodyGyroJerk-XYZ  tBodyAccMag  tGravityAccMag  tBodyAccJerkMag  tBodyGyroMag  tBodyGyroJerkMag  fBodyAcc-XYZ  fBodyAccJerk-XYZ  fBodyGyro-XYZ  fBodyAccMag  fBodyAccJerkMag  fBodyGyroMag  fBodyGyroJerkMag  Variables assigned:  mean()  std()  mad()  max()  min()  sma()  energy()  iqr()  entropy()  arCoeff()  correlation()  maxInds()  meanFreq()  skewness()  kurtosis()  bandsEnergy()  angle() |
| Features\_info.txt | NA | NA | Description of features variables | NA |
| subject\_test.txt | 1 | 2947 | Subject IDs for (9) randomly assigned test subjects; identifies subject who performed activity for each sample | 2, 4, 9, 10, 12, 13, 18, 20, 24 |
| X\_test.txt | 561 | 2947 | Training set | [continuous] |
| y\_test.txt | 1 | 2947 | Activity indexes | 1, 2, 3, 4, 5, 6 |
| subject\_train.txt | 1 | 7352 | Subject IDs for (21) randomly assigned test subjects; identifies subject who performed activity for each sample | 1, 3, 5, 6, 7, 8, 11, 14, 15, 16, 17, 19, 21, 22, 23, 25, 26, 27, 28, 29, 30 |
| X\_train.txt | 561 | 7352 | Training set | [continuous] |
| y\_train.txt | 1 | 7352 | Class labels | 1, 2, 3, 4, 5, 6 |

TIDY DATA SET #1 CONSTRUCTION:

The composite data set was constructed as follows:

1. Combine training and test data sets into one data set (10299 records x 561 variables), comprised of:

* train/X\_train.txt (7352 records x 561 variables)
* test/X\_test.txt (2947 records x 561 variables)

1. Extract feature names from file and set as column names for data set:
   * features.txt (561 records x 2 variables)
2. Add features as header names to columns in data set(10299 records x 561 variables + headers)
3. Extract only Mean and Standard Deviation (std) measurements from the dataset (10299 records x 66 variables)
4. Combine training and test activity label files into one list and assign name ActivityIndex to list (10299 records), comprised of:

* train/y\_train.txt (7352 records)
* test/y\_test.txt (2947 records)

1. Combine subject IDS for training and test sets into one list and assign name SubjectID to list (10299 records), comprised of:

* train/subject\_train.txt (7352 records)
* test/subject\_test.txt (2947 records)

1. Append ActivityIndex and SubjectID to dataset (10299 records x 68 variables)
2. Merge activity labels with dataset based on ActivityIndex (10299 records x 69 variables), comprised of:
   * activity\_labels.txt (6 records x 2 variables)
3. Delete observations that do not appear to contain meaningful data per descriptors in the UCI MLR README file, specifically:
   * Those for the variable “BodyBody”
4. Write data to file Smartphones\_inertial\_sensors.csv

TIDY DATA SET #2 CONSTRUCTION:

The second tidy data set was constructed as follows:

1. Melt the composite data set to consolidate column variables.
2. Calculate means by subject and activity.
3. Write data to file Activity\_means.csv

DATA DICTIONARY

|  |  |  |
| --- | --- | --- |
| VARIABLE | DESCRIPTION | VALUE OR EXPLANATION |
| SubjectID | Subject identification numbers | 1-30 (2, 4, 9, 10, 12, 13, 18, 20, 24 (n=9) for test data; 1, 3, 5, 6, 7, 8, 11, 14, 15, 16, 17, 19, 21, 22, 23, 25, 26, 27, 28, 29, 30 (n=21) for training data) |
| ActivityIndex | Index number for specific activites | 1-6 |
| ActivityName | Activities corresponding with ActivityIndex | WALKING  WALKING\_UPSTAIRS  WALKING\_DOWNSTAIRS  SITTING  STANDING  LAYING |
| Time.Body.Accelerometer-mean()-X | Mean body linear acceleration,x-dimension, time domain | continuous |
| Time.Body.Accelerometer-mean()-Y | Mean body linear acceleration,x-dimension, time domain | continuous |
| Time.Body.Accelerometer-mean()-Z | Mean body linear acceleration,x-dimension, time domain | continuous |
| Time.Body.Accelerometer-std()-X | Standard deviation body linear acceleration,x-dimension, time domain | continuous |
| Time.Body.Accelerometer-std()-Y | Standard deviation body linear acceleration,y-dimension, time domain | continuous |
| Time.Body.Accelerometer-std()-Z | Standard deviation body linear acceleration,z-dimension, time domain | continuous |
| Time.Gravity.Accelerometer-mean()-X | Mean gravity linear acceleration,x-dimension, time domain | continuous |
| Time.Gravity.Accelerometer-mean()-Y | Mean gravity linear acceleration,x-dimension, time domain | continuous |
| Time.Gravity.Accelerometer-mean()-Z | Mean gravity linear acceleration,x-dimension, time domain | continuous |
| Time.Gravity.Accelerometer-std()-X | Standard deviation gravity linear acceleration,x-dimension, time domain | continuous |
| Time.Gravity.Accelerometer-std()-Y | Standard deviation gravity linear acceleration,x-dimension, time domain | continuous |
| Time.Gravity.Accelerometer-std()-Z | Standard deviation gravity linear acceleration,x-dimension, time domain | continuous |
| Time.Body.Accelerometer-Jerk-mean()-X | Mean body linear acceleration jerk signal,x-dimension, time domain | continuous |
| Time.Body.Accelerometer-Jerk-mean()-Y | Mean body linear acceleration jerk signal,y-dimension, time domain | continuous |
| Time.Body.Accelerometer-Jerk-mean()-Z | Mean body linear acceleration jerk signal,z-dimension, time domain | continuous |
| Time.Body.Accelerometer-Jerk-std()-X | Standard deviation body linear acceleration jerk signal,x-dimension, time domain | continuous |
| Time.Body.Accelerometer-Jerk-std()-Y | Standard deviation body linear acceleration jerk signal,y-dimension, time domain | continuous |
| Time.Body.Accelerometer-Jerk-std()-Z | Standard deviation body linear acceleration jerk signal,z-dimension, time domain | continuous |
| Time.Body.Gyroscope-mean()-X | Mean body gyroscope reading,x-dimension, time domain | continuous |
| Time.Body.Gyroscope-mean()-Y | Mean body gyroscope reading,y-dimension, time domain | continuous |
| Time.Body.Gyroscope-mean()-Z | Mean body gyroscope reading,z-dimension, time domain | continuous |
| Time.Body.Gyroscope-std()-X | Standard deviation body gyroscope reading,x-dimension, time domain | continuous |
| Time.Body.Gyroscope-std()-Y | Standard deviation body gyroscope reading,y-dimension, time domain | continuous |
| Time.Body.Gyroscope-std()-Z | Standard deviation body gyroscope reading,z-dimension, time domain | continuous |
| Time.Body.Gyroscope-Jerk-mean()-X | Mean body gyroscope reading,jerk signal, x-dimension, time domain | continuous |
| Time.Body.Gyroscope-Jerk-mean()-Y | Mean body gyroscope reading,jerk signal, y-dimension, time domain | continuous |
| Time.Body.Gyroscope-Jerk-mean()-Z | Mean body gyroscope reading,jerk signal, z-dimension, time domain | continuous |
| Time.Body.Gyroscope-Jerk-std()-X | Standard deviation body gyroscope reading,jerk signal, x-dimension, time domain | continuous |
| Time.Body.Gyroscope-Jerk-std()-Y | Standard deviation body gyroscope reading,jerk signal, y-dimension, time domain | continuous |
| Time.Body.Gyroscope-Jerk-std()-Z | Standard deviation body gyroscope reading,jerk signal, z-dimension, time domain | continuous |
| Time.Body.Accelerometer-Mag-mean() | Mean body linear acceleration magnitude, time domain | continuous |
| Time.Body.Accelerometer-Mag-std() | Standard deviation body linear acceleration magnitude, time domain | continuous |
| Time.Gravity.Accelerometer-Mag-mean() | Mean gravity linear acceleration magnitude, time domain | continuous |
| Time.Gravity.Accelerometer-Mag-std() | Standard deviation gravity linear acceleration magnitude, time domain | continuous |
| Time.Body.Accelerometer-JerkMag-mean() | Mean body linear acceleration jerk signal magnitude, time domain | continuous |
| Time.Body.Accelerometer-JerkMag-std() | Standard deviation body linear acceleration jerk signal magnitude, time domain | continuous |
| Time.Body.Gyroscope-Mag-mean() | Mean body gyroscope reading magnitude, time domain | continuous |
| Time.Body.Gyroscope-Mag-std() | Standard deviation body gyroscope reading magnitude, time domain | continuous |
| Time.Body.Gyroscope-JerkMag-mean() | Mean body gyroscope reading, jerk signal magnitude, time domain | continuous |
| Time.Body.Gyroscope-JerkMag-std() | Standard deviation body gyroscope reading, jerk signal magnitude, time domain | continuous |
| Freq.Body.Accelerometer-mean()-X | Mean body linear acceleration, x-dimension, frequency domain | continuous |
| Freq.Body.Accelerometer-mean()-Y | Mean body linear acceleration, y-dimension, frequency domain | continuous |
| Freq.Body.Accelerometer-mean()-Z | Mean body linear acceleration, z-dimension, frequency domain | continuous |
| Freq.Body.Accelerometer-std()-X | Standard deviation body linear acceleration, x-dimension, frequency domain | continuous |
| Freq.Body.Accelerometer-std()-Y | Standard deviation body linear acceleration, y-dimension, frequency domain | continuous |
| Freq.Body.Accelerometer-std()-Z | Standard deviation body linear acceleration, z-dimension, frequency domain | continuous |
| Freq.Body.Accelerometer-Jerk-mean()-X | Mean body linear acceleration jerk signal,x-direction, frequency domain | continuous |
| Freq.Body.Accelerometer-Jerk-mean()-Y | Mean body linear acceleration jerk signal,y-direction, frequency domain | continuous |
| Freq.Body.Accelerometer-Jerk-mean()-Z | Mean body linear acceleration jerk signal,z-direction, frequency domain | continuous |
| Freq.Body.Accelerometer-Jerk-std()-X | Standard deviation body linear acceleration jerk signal,x-direction, frequency domain | continuous |
| Freq.Body.Accelerometer-Jerk-std()-Y | Standard deviation body linear acceleration jerk signal,y-direction, frequency domain | continuous |
| Freq.Body.Accelerometer-Jerk-std()-Z | Standard deviation body linear acceleration jerk signal,z-direction, frequency domain | continuous |
| Freq.Body.Gyroscope-mean()-X | Mean body gyroscope reading, x-dimension, frequency domain | continuous |
| Freq.Body.Gyroscope-mean()-Y | Mean body gyroscope reading, y-dimension, frequency domain | continuous |
| Freq.Body.Gyroscope-mean()-Z | Mean body gyroscope reading, z-dimension, frequency domain | continuous |
| Freq.Body.Gyroscope-std()-X | Standard deviation body gyroscope reading, x-dimension, frequency domain | continuous |
| Freq.Body.Gyroscope-std()-Y | Standard deviation body gyroscope reading, y-dimension, frequency domain | continuous |
| Freq.Body.Gyroscope-std()-Z | Standard deviation body gyroscope reading, z-dimension, frequency domain | continuous |
| Freq.Body.Accelerometer-Mag-mean() | Mean body linear acceleration jerk signal magnitude, frequency domain | continuous |
| Freq.Body.Accelerometer-Mag-std() | Standard deviation body linear acceleration jerk signal magnitude, frequency domain | continuous |