Human Activity Recognition Using Smartphones Data Set
Download: Data Folder, Data Set Description

Abstract: Human Activity Recognition database built from the recordings of 30 subjects performing activities of daily living (ADL) while carrying a waist-mounted smartphone with embedded inertial sensors.

Data Set Characteristics:	Multivariate, Time-Series	**Number of Instances:**	10299	**Area:**	Computer
Attribute Characteristics:	N/A	**Number of Attributes:**	561	**Date Donated**	2012-12-10
Associated Tasks:	Classification, Clustering	**Missing Values?**	N/A	**Number of Web Hits:**	530633

Source:

Jorge L. Reyes-Ortiz(1,2), Davide Anguita(1), Alessandro Ghio(1), Luca Oneto(1) and Xavier Parra(2)
1 - Smartlab - Non-Linear Complex Systems Laboratory
DITEN - Università degli Studi di Genova, Genoa (I-16145), Italy.
2 - CETpD - Technical Research Centre for Dependency Care and Autonomous Living
Universitat Politècnica de Catalunya (BarcelonaTech). Vilanova i la Geltrú (08800), Spain
activityrecognition '@' smartlab.ws

Updated Codebook:

Original Codebook credited to Jorge L. Reyes-Ortiz(1,2), Davide Anguita(1), Alessandro Ghio(1), Luca Oneto(1) and Xavier Parra(2) 1 - Smartlab - Non-Linear Complex Systems Laboratory.

Updates made by Kimberley Kirk for Data Science Specialization through Johns Hopkins University.

Data Set Information:

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.

Check the README.txt file for further details about this dataset.

A video of the experiment including an example of the 6 recorded activities with one of the participants can be seen in the following link: [Web Link]

An updated version of this dataset can be found at [Web Link]. It includes labels of postural transitions between activities and also the full raw inertial signals instead of the ones pre-processed into windows.

Attribute Information:

For each record in the dataset it is provided:
Triaxial acceleration from the accelerometer (total acceleration) and the estimated body acceleration
UPDATED: for mean, mean frequency, and standard deviation only

Triaxial Angular velocity from the gyroscope.

UPDATED: for mean, mean frequency, and standard deviation only
86 “feature” variables with time and frequency domain variables for mean, mean frequency, and standard deviation only.

Its activity label, labeled "activities" variable

An identifier of the subject who carried out the experiment. Labeled "subject" variable

Data variables include:

"subject”
"activities"
"T_XAxis_BodyAccelerometer_Mean”
"T_YAxis_BodyAccelerometer_Mean"
"T_ZAxis_BodyAccelerometer_Mean”
"T_XAxis_GravityAccelerometer_Mean"
"T_YAxis_GravityAccelerometer_Mean”
"T_ZAxis_GravityAccelerometer_Mean"
"T_XAxis_BodyAccelerometerJerk_Mean"
"T_YAxis_BodyAccelerometerJerk_Mean"
"T_ZAxis_BodyAccelerometerJerk_Mean"
"T_XAxis_BodyGyroscope_Mean"
"T_YAxis_BodyGyroscope_Mean"
"T_ZAxis_BodyGyroscope_Mean"
"T_XAxis_BodyGyroscopeJerk_Mean"
"T_YAxis_BodyGyroscopeJerk_Mean"
"T_ZAxis_BodyGyroscopeJerk_Mean"
"T_BodyAccelerometerMagnitude_Mean"
"T_GravityAccelerometerMagnitude_Mean"
"T_BodyAccelerometerJerkMagnitude_Mean"
"T_BodyGyroscopeMagnitude_Mean"
"T_BodyGyroscopeJerkMagnitude_Mean"
"F_XAxis_BodyAccelerometer_Mean"

"F_YAxis_BodyAccelerometer_Mean"
"F_ZAxis_BodyAccelerometer_Mean"
"F_XAxis_BodyAccelerometer_MeanFrequency"
"F_YAxis_BodyAccelerometer_MeanFrequency"
"F_ZAxis_BodyAccelerometer_MeanFrequency"
"F_XAxis_BodyAccelerometerJerk_Mean"
"F_YAxis_BodyAccelerometerJerk_Mean"
"F_ZAxis_BodyAccelerometerJerk_Mean"
"F_XAxis_BodyAccelerometerJerk_MeanFrequency"
"F_YAxis_BodyAccelerometerJerk_MeanFrequency"
"F_ZAxis_BodyAccelerometerJerk_MeanFrequency"
"F_XAxis_BodyGyroscope_Mean"
"F_YAxis_BodyGyroscope_Mean"
"F_ZAxis_BodyGyroscope_Mean"
"F_XAxis_BodyGyroscope_MeanFrequency"
"F_YAxis_BodyGyroscope_MeanFrequency"
"F_ZAxis_BodyGyroscope_MeanFrequency"
"F_BodyAccelerometerMagnitude_Mean"
"F_BodyAccelerometerMagnitude_MeanFrequency"
"F_BodyBodyAccelerometerJerkMagnitude_Mean"
"F_BodyBodyAccelerometerJerkMagnitude_MeanFrequency"
"F_BodyBodyGyroscopeMagnitude_Mean"
"F_BodyBodyGyroscopeMagnitude_MeanFrequency"
"F_BodyBodyGyroscopeJerkMagnitude_Mean"
"F_BodyBodyGyroscopeJerkMagnitude_MeanFrequency"
"Angle_TimeDomainSignal_BodyAccelerometerMean_Gravity"
"Angle_TimeDomainSignal_BodyAccelerometerJerkMean_GravityMean"
"Angle_TimeDomainSignal_BodyGyroscopeMean_GravityMean"
"Angle_TimeDomainSignal_BodBodyGyroscopeJerkMean_GravityMean"
"Angle_XAxis_GravityMean"
"Angle_YAxis_GravityMean"
"Angle_ZAxis_GravityMean"
"T_XAxis_BodyAccelerometer_SD"

"T_YAxis_BodyAccelerometer_SD"
"T_ZAxis_BodyAccelerometer_SD"
"T_XAxis_GravityAccelerometer_SD"
"T_YAxis_GravityAccelerometer_SD"
"T_ZAxis_GravityAccelerometer_SD"
"T_XAxis_BodyAccelerometerJerk_SD"
"T_YAxis_BodyAccelerometerJerk_SD"
"T_ZAxis_BodyAccelerometerJerk_SD"
"T_XAxis_BodyGyroscope_SD"
"T_YAxis_BodyGyroscope_SD"
"T_ZAxis_BodyGyroscope_SD"
"T_XAxis_BodyGyroscopeJerk_SD"
"T_YAxis_BodyGyroscopeJerk_SD"
"T_ZAxis_BodyGyroscopeJerk_SD"
"T_BodyAccelerometerMagnitude_SD"
"T_GravityAccelerometerMagnitude_SD"
"T_BodyAccelerometerJerkMagnitude_SD"
"T_BodyGyroscopeMagnitude_SD"
"T_BodyGyroscopeJerkMagnitude_SD"
"F_XAxis_BodyAccelerometer_SD"
"F_YAxis_BodyAccelerometer_SD"
"F_ZAxis_BodyAccelerometer_SD"
"F_XAxis_BodyAccelerometerJerk_SD"
"F_YAxis_BodyAccelerometerJerk_SD"
"F_ZAxis_BodyAccelerometerJerk_SD"
"F_XAxis_BodyGyroscope_SD"
"F_YAxis_BodyGyroscope_SD"
"F_ZAxis_BodyGyroscope_SD"
"F_BodyAccelerometerMagnitude_SD"
"F_BodyBodyAccelerometerJerkMagnitude_SD"

Data Summaries Calculated:

for each subject's activity, the average is calculated for all 86 “features” variables

Units for Data Summaries and Variables:

"subject" variable is in one human being units

"activities" variable is in activity units WALKING, WALKING_UPSTAIRS, WALKING_DOWNSTAIRS, SITTING, STANDING, LAYING units

86 variables that are in Hz and are normalized and bounded within [-1,1]; These are summarized variables that show the average

Data Transformations:

Zip file “getdata_projectfiles_UCI HAR Dataset” was downloaded from host website

Set file path to save download to

Set url for download

Download file and save to working directory

“y_test" and "y_train" were converted into Factor with 6 levels

“x_train” was converted into numeric, non scientific notation

"subject_train" was converted into Factor with 21 levels “subject_test” was converted into Factor with 9 levels

“x_test” was converted into numeric, non scientific notation

“features” was converted into character vector and used as headings for “x_train” and “x_test”

“subject_train” and “subject_test” were bound to “x_train” and “x_test” respectively

“y_test” and “y_train” were bound together by column “x_test” and “x_train” respectively

“x_test” and “x_train” were bound together by row

Check for dplyr package, prompt user to download if no required package, load dplyr library

Columns were subset for “mean” and “standard deviation” for the “x_test” and “x_train” combined datasets

Average was calculated (using the “mean” function) for all numeric variables in the “x_test” and “x_train” combined datasets grouped by “subject” and “activities” variables (aka final “tidy dataset”)

Variable names were assigned to the final “tidy dataset”

Relevant Papers:

Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra and Jorge L. Reyes-Ortiz. Human Activity Recognition on Smartphones using a Multiclass Hardware-Friendly Support Vector Machine. International Workshop of Ambient Assisted Living (IWAAL 2012). Vitoria-Gasteiz, Spain. Dec 2012

Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra, Jorge L. Reyes-Ortiz. Energy Efficient Smartphone-Based Activity Recognition using Fixed-Point Arithmetic. Journal of Universal Computer Science. Special Issue in Ambient Assisted Living: Home Care. Volume 19, Issue 9. May 2013

Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra and Jorge L. Reyes-Ortiz. Human Activity Recognition on Smartphones using a Multiclass Hardware-Friendly Support Vector Machine. 4th International Workshop of

Ambient Assited Living, IWAAL 2012, Vitoria-Gasteiz, Spain, December 3-5, 2012. Proceedings. Lecture Notes in Computer Science 2012, pp 216-223.

Jorge Luis Reyes-Ortiz, Alessandro Ghio, Xavier Parra-Llanas, Davide Anguita, Joan Cabestany, Andreu Català. Human Activity and Motion Disorder Recognition: Towards Smarter Interactive Cognitive Environments. 21th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN 2013. Bruges, Belgium 24-26 April 2013.

Citation Request:

Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra and Jorge L. Reyes-Ortiz. A Public Domain Dataset for Human Activity Recognition Using Smartphones. 21th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN 2013. Bruges, Belgium 24-26 April 2013.