|  |  |
| --- | --- |
| |  | | --- | | **Smartphone-Based Recognition of Human Activities and Postural Transitions Data Set** | |

**Abstract**: The challenge in this project is creating a ML algorithm for activity recognition using a set of sensor data. The dataset given is built from the recordings of 30 subjects performing basic activities and postural transitions while carrying a waist-mounted smartphone with embedded inertial sensors. The characteristics of the data is summarized below.

|  |  |  |  |
| --- | --- | --- | --- |
| **Data Set Characteristics:** | Multivariate, Time-Series | **Number of Instances:** | 10929 |
| **Attribute Characteristics:** | Real | **Number of Attributes:** | 561 |
| **Associated Tasks:** | Classification | **Missing Values?** | N/A |

**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.

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 561 features was obtained by calculating variables from the time and frequency domain. See 'features\_info.txt' for more details.

**Attribute Information:**

The dataset is then divided in two parts and they can be used separately.

1. Inertial sensor data
   * Raw triaxial signals from the accelerometer and gyroscope of all the trials with with participants.
   * The labels of all the performed activities.
2. Records of activity windows. Each one composed of:
   * A 561-feature vector with time and frequency domain variables.
   * Its associated activity label.
   * An identifier of the subject who carried out the experiment.

The dataset includes the following files:   
========================================= 

* + 'README.txt'
  + ‘RawData/acc\_expXX\_userYY.txt’: The raw triaxial acceleration signal for the experiment number XX and associated to the user number YY. Every row is one acceleration sample (three axis) captured at a frequency of 50Hz.
  + ‘RawData/gyro\_expXX\_userYY.txt’: The raw triaxial angular speed signal for the experiment number XX and associated to the user number YY. Every row is one angular velocity sample (three axis) captured at a frequency of 50Hz.
  + ‘RawData/labels.txtActivity’: labels available for the dataset (1 per row).   
    Column 1: experiment number ID,   
    Column 2: user number ID,   
    Column 3: activity number ID   
    Column 4: Label start point (in number of signal log samples (recorded at 50Hz))   
    Column 5: Label end point (in number of signal log samples)
  + 'features\_info.txt': Shows information about the variables used on the feature vector.
  + 'features.txt': List of all features.
  + 'activity\_labels.txt': Links the activity ID with their activity name.
  + ‘Train/X\_train.txt’: Training set.
  + ‘Train/y\_train.txt’: Training labels.
  + ‘Test/y\_test.txt’: Test set.
  + ‘Test/y\_test.txt’: Test labels.
  + ‘Train/subject\_id\_train.txt’: Each row identifies the subject who performed the activity for each window sample. Its range is from 1 to 30.
  + ‘Test/subject\_id\_test.txt’: Each row identifies the subject who performed the activity for each window sample. Its range is from 1 to 30.

**Deliverables:**

1. The python code used to implement your solution
2. A **short** description (in a word document) of your solution which justifies your reasoning and summarizes your result

**N.B. You will be evaluated based on your coding and your machine learning skills**