**Taru Tak 04/28/2018**

**MIS 637**

**Project Proposal**

**Activity Recognition system based on Multi-sensor Data Fusion**

**Problem Statement**

Classify human activities into bending, cycling, lying down, sitting, standing, walking based on data from a wireless sensor network worn by an actor performing these activities.

**Objective**

Multi-label classification of user activities based on sensor data using data mining techniques.

**Methodology**

The methodology to be used for this project will be CRISP-DM and the data mining techniques used are

1. Logistic Regression
2. Naïve bayes classification
3. Decision trees
4. SVM classifiers

* To evaluate the performance of the algorithms using statistical measures like accuracy, confusion matrix, precision and recall ROC AUC.
* Hyper tuning of parameter to obtain the best performance.

**About the Dataset**

Dataset Link: <https://archive.ics.uci.edu/ml/datasets/Activity+Recognition+system+based+on+Multisensor+data+fusion+(AReM)>

|  |  |
| --- | --- |
| **Abstract**: This dataset contains temporal data from a Wireless Sensor Network worn by an actor performing the activities: bending, cycling, lying down, sitting, standing, and walking. |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Data Set Characteristics:** | Multivariate, Sequential, Time-Series | **Number of Instances:** | 42240 | **Area:** | Computer |
| **Attribute Characteristics:** | Real | **Number of Attributes:** | 6 | **Date Donated** | 2016-05-18 |
| **Associated Tasks:** | Classification | **Missing Values?** | N/A | **Number of Web Hits:** | 29610 |

**Data Set Information:**

This dataset represents a real-life benchmark in the area of Activity Recognition applications.  
The classification tasks consist in predicting the activity performed by the user from time-series generated by a Wireless Sensor Network (WSN). From the raw data we extract time-domain features to compress the time series and slightly remove noise and correlations. The three couples of WSN nodes (i.e. Chest-Right Ankle, Chest-Left Ankle, Right Ankle-Left Ankle). The features include the mean value and standard deviation for each reciprocal RSS reading from worn WSN sensors.

The dataset contains 7 folders that represent seven types of activities. In each folder, there are multiples each of which represents an instant of a human performing an activity. Each contain is 6 time series collected from activities of the same person. There are 88 instances in the dataset, each of which contains 6 time series and each time series has 480 consecutive values.

**Sample data:**

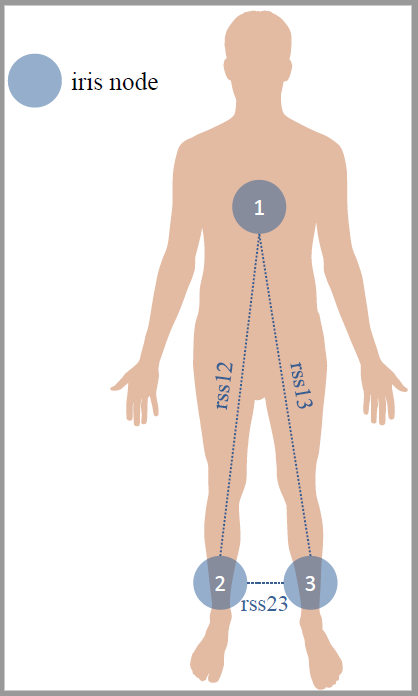
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| # Task: bending1 | |  |  |  |  |  |
| # Frequency (Hz): 20 | |  |  |  |  |  |
| # Clock (millisecond): 250 | | |  |  |  |  |
| # Duration (seconds): 120 | | |  |  |  |  |
| # Columns: time | avg\_rss12 | var\_rss12 | avg\_rss13 | var\_rss13 | avg\_rss23 | var\_rss23 |
| 0 | 39.25 | 0.43 | 22.75 | 0.43 | 33.75 | 1.3 |
| 250 | 39.25 | 0.43 | 23 | 0 | 33 | 0 |
| 500 | 39.25 | 0.43 | 23.25 | 0.43 | 33 | 0 |
| 750 | 39.5 | 0.5 | 23 | 0.71 | 33 | 0 |
| 1000 | 39.5 | 0.5 | 24 | 0 | 33 | 0 |
| 1250 | 39.25 | 0.43 | 24 | 0 | 33 | 0 |
| 1500 | 39.25 | 0.43 | 24 | 0 | 33 | 0 |
| 1750 | 39 | 0 | 23.75 | 0.43 | 33 | 0 |
| 2000 | 39.5 | 0.5 | 24 | 0 | 33 | 0 |
| 2250 | 39.5 | 0.5 | 23 | 0 | 33 | 0 |
| 2500 | 39.5 | 0.5 | 23.25 | 0.43 | 33 | 0 |
| 2750 | 39.5 | 0.5 | 23.5 | 0.5 | 32.75 | 0.43 |
| 3000 | 39.5 | 0.5 | 23.75 | 0.43 | 32.5 | 0.5 |
| 3250 | 39.67 | 0.47 | 23.75 | 0.43 | 33 | 0 |
| 3500 | 39.5 | 0.5 | 24 | 0 | 33 | 0 |
| 3750 | 39.5 | 0.5 | 23.25 | 0.43 | 33 | 0 |
| 4000 | 39.5 | 0.5 | 22.5 | 0.5 | 33 | 0 |
| 4250 | 39.5 | 0.5 | 22 | 0.71 | 33 | 0 |
| 4500 | 40.25 | 0.83 | 21 | 0 | 33 | 0 |
| 4750 | 40.5 | 0.5 | 18.67 | 1.7 | 33 | 0 |
| 5000 | 40.67 | 0.47 | 15.5 | 0.87 | 33 | 0 |
| 5250 | 40.5 | 0.5 | 15 | 0 | 33 | 0 |
| 5500 | 40.5 | 0.5 | 15 | 0 | 33 | 0 |
| 5750 | 40.33 | 0.47 | 16 | 0 | 33 | 0 |

**Attribute Information:**

For each sequence, data is provided in comma separated value (csv) format.   
- Input data:   
Input RSS streams are provided in files named datasetID.csv, where ID is the progressive numeric sequence ID for each repetition of the activity performed.   
In each file, each row corresponds to a time step measurement (in temporal order) and contains the following information:   
avg\_rss12, var\_rss12, avg\_rss13, var\_rss13, avg\_rss23, var\_rss23   
where avg and var are the mean and variance values over 250 ms of data, respectively.   
- Target data:   
Target data is provided as the containing folder name.

For each activity, we have the following parameters:   
# Frequency (Hz): 20   
# Clock (millisecond): 250   
# Total duration (seconds): 120

**Sensor placement**



**Citation**

F. Palumbo, C. Gallicchio, R. Pucci and A. Micheli, Human activity recognition using multisensor data fusion based on Reservoir Computing, Journal of Ambient Intelligence and Smart Environments, 2016, 8 (2), pp. 87-107.