Dataset Description

The dataset presented in this project is the “Activities of Daily Living”, which can be found in the Machine Learning Repository (UCI). Available in:

<https://archive.ics.uci.edu/ml/datasets/Activities+of+Daily+Living+(ADLs)+Recognition+Using+Binary+Sensors>

The dataset includes sensors located in a house, which are activated every time a person moves. The sensors are installed in specific location in the house. This is a 4 rooms house, and the person’s moves have been labeled for 14 days. There is a total of 5 sensor types (PIR, Flush, Electric etc.) which are located in 12 different locations in the house (Seat, Bed, Microwave, Shower etc. Furthermore, the dataset includes start and stop date and time both for the sensors and for the activities.

The labeled activities are Breakfast, Showering, Toileting, Sleeping etc. Α brief descripition of the different locations, types, places and activities of the dataset is given in the table below:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| No. | Location | Type | Place | Activity |
| 1 | Bed | Pressure | Bedroom | Sleeping |
| 2 | Cabinet | Magnetic | Bathroom | Toileting |
| 3 | Basin | PIR | Kitchen | Showering |
| 4 | Toilet | Flush | Living Room | Breakfast |
| 5 | Shower | Electric | Entrance | Grooming |
| 6 | Fridge |  |  | Spare\_Time/TV |
| 7 | Cupboard |  |  | Leaving |
| 8 | Toaster |  |  | Lunck |
| 9 | Cooktop |  |  | Snack |
| 10 | Microwave |  |  |  |
| 11 | Seat |  |  |  |
| 12 | Maindoor |  |  |  |

In the drawing below, one can see the top view of the house. In this drawing a more comprehensive view of the sensors’ distribution can be obtained.



Data Storytelling (pt.1) : What’s the time bro?

The purpose of the project, is the implementation of one (or more) machine learning algorithm(s) to predict the human activity depending on each sensor’s activation. The basic idea is to find a way to combine the human activity and the censor activation. The key to unlock this mystery hides in the “time” attribute. Comparing the “time” columns both for the sensors activation and for the human activity one can observe that for many cases during a human activity there is more than one sensor activated. So, all there is to be investigated, is the possible relationship between the start/end time for each case.

The most important observation was that the activation of a specific type of sensor in a specific place in the house meant one and only specific activity. For example, the “Pressure” type activation which is located in the bedroom meant “Sleeping” .Bearing this in mind, a scenario was built, based on two inequalities. With the guidance of this scenario, an activity label column was created, in the sensor activation dataset. The labels for each activity were numbers ranging from 1-9 exactly as they are displayed in the table above.

**Case1:**

Sensor activation Start time >= Activity Start Time

AND

Sensor activation Start time <= Activity End Time

Act: Activity values are registered in the censor dataset.

**Case 2:**

Sensor activation End time >= Activity Start Time

AND

Sensor activation End time <=Activity End Time

Act: Activity values are registered in the censor dataset.

**Case 3:**

If case 1 AND case 2 Are False🡪 examine whether the previous activity was one of the unique cases mentioned above.

Act: Activity values are registered in the censor dataset.

**Case 4:**

If case1,2,3 are all false, 🡪 register the previous index human activity in the censor dataset.

Data Storytelling (Part.2) A better view

The final dataset created was like this (first 10 rows):

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 0 | Location | Type | Place | St.Time | End.Time | Act.Labels | Activity |
| 1 | Bed | Pressure | Bedroom | (11/11/28 02:27:59) | (11/11/28 10:18:11) | 1 | Sleeping |
| 2 | Cabinet | Magnetic | Bathroom | (11/11/28 10:21:24) | (11/11/28 10:21:31) | 2 | Toileting |
| 3 | Basin | PIR | Bathroom | (11/11/28 10:21:44) | (11/11/28 10:23:31) | 2 | Toileting |
| 4 | Toilet | Flush | Bathroom | (11/11/28 10:23:02) | (11/11/28 10:23:36) | 2 | Toileting |
| 5 | Shower | PIR | Bathroom | (11/11/28 10:25:44) | (11/11/28 10:32:06) | 3 | Showering |
| 6 | Fridge | Magnetic | Kitchen | (11/11/28 10:34:23) | (11/11/28 10:34:41) | 4 | Breakfast |
| 7 | Cupboard | Magnetic | Kitchen | (11/11/28 10:34:44) | (11/11/28 10:37:17) | 4 | Breakfast |
| 8 | Toaster | Electric | Kitchen | (11/11/28 10:38:00) | (11/11/28 10:42:41) | 4 | Breakfast |
| 9 | Fridge | Magnetic | Kitchen | (11/11/28 10:38:33) | (11/11/28 10:38:40) | 4 | Breakfast |
| 10 | Cupboard | Magnetic | Kitchen | (11/11/28 10:41:29) | (11/11/28 10:41:36) | 4 | Breakfast |

Now it seems that things are made a little more obvious for this project. This is a multi-label classification problem where the Act. Label column is the class to be predicted after the implementation of the classifiers. Furthermore, one can see the difference among the sensor activation and the human activity (for more than one censor activations, human activity remains the same).