Student ID: s3831858
Student Name: Varsha Chandrasekhar Bendre

I certify that this is all my own original work. If I took any parts from elsewhere, then they were non-essential parts of the assignment, and they are clearly attributed in my submission. I will show I agree to this honor code by typing "Yes": *Yes*.

Activity Recognition from Single Chest-Mounted Accelerometer Data Set

Masters in AI student, RMIT University
s3831858@student.rmit.edu.au
Date of Report: 8 June 2020

Table of Contents

1. An abstract/executive summary

The aim of the report was to identify the activities for research purposes based on the un-calibrated accelerometer sensor data obtained and resist the activities that might be harmful if continued. The Sensors are mounted on the User's chest and then record the sensor readings. Overall, the results indicate that Task 1 (Working at Computer) is the most performed activity by all the Users. The report concludes that people spend most of their time at the Computer and not on physical activities, which is not a good practice for anyone. It is recommended that people spend more time on other activities which involve physical exercise along with Task 1, as it is vital to have good physical health.

2. Introduction

The chest-mounted accelerometer, which was custom developed, collects the data from the person of interest at 52 Hz i.e., 52 observations per second. Accelerometers are sensors that calculate a moving body's acceleration and also detect the frequency and intensity of movements in the human body. The accelerometers worn on the body have shown good accuracies for studying the classification of activities such as standing, sitting, walking, and others. As these sensors are compact with a low-power requirement and also has the capability to store the data that is related to the motion of the people. This report will discuss the activities' identification and classification based on the sensor data and avoid the activities that might be harmful if continued for a more extended period.(Brownlee 2018)

Figure 1: (Brownlee 2018) Chest mounted Accelerometer

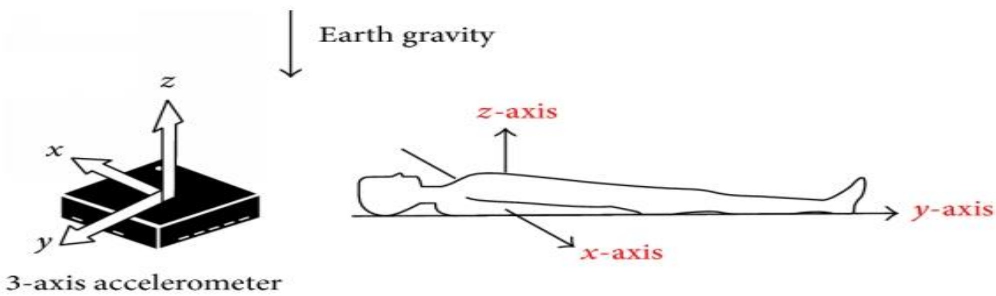


Figure 2:(Tadi, Koivisto, Pänkäälä, Paasio, 2014) 3-axis accelerometer

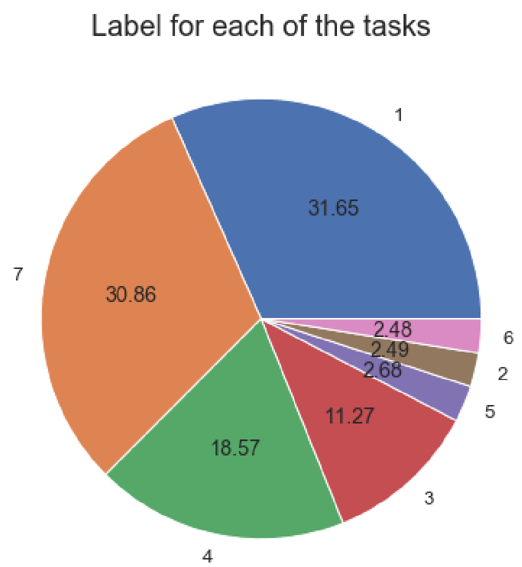
3. Methodology

The data was collected from 15 Users whose age was between 27 and 35. The 7 activities are Task 1: Working at Computer, Task 2: Standing Up, Walking and Going up/downstairs, Task 3: Standing, Task 4: Walking, Task 5: Going Up/Downstairs, Task 6: Walking and Talking with Someone , Task 7: Talking while Standing . As there is no regular method to relate the data from the sensor to the specific activities, it is a difficult problem. The activities were labelled with the help of people by asking them to note down the order of their activities and restart their accelerometer. And when the sensor is restarted, serial number is given to the data. After starting the sensor, User can do the task. It takes about 2 minutes for the system to boot and it starts collecting the data while the User is performing. The sensor can be stopped by pressing the start button again. The data volume is vast as there are a few hundreds of observations per second; hence it is also a computationally challenging problem. The variables chosen are acceleration in the x-axis, y-axis, z-axis, User ID and the Tasks (Labels in dataset). The Decision tree classifier and K-Nearest Neighbours models are used on the dataset with hill-climbing technique and parameter tuning effect (Casale, Radeva, & Pujol 2011).

4. Results

<u>4.1.Results from column exploration</u>

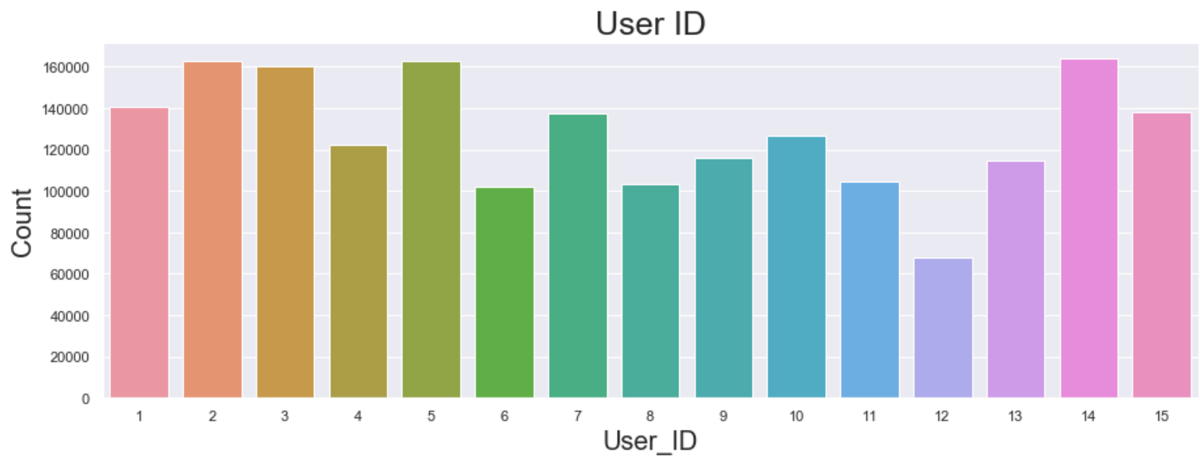
1. Label column



a) People have spent most of the time on task 1 i.e. Working at Computer and the least time on task 6 i.e. Walking and Talking with Someone.
b) Task 2,6,5 are the least performed actions. which is around 2.5%
c) Task 1,7 are the most performed actions, which is around 30%.

2. User ID column

a) Maximum amount of data is available from the Users 2,3,5,14 which is around 160000 data entries.
b) Minimum amount of data is available from the User 12.



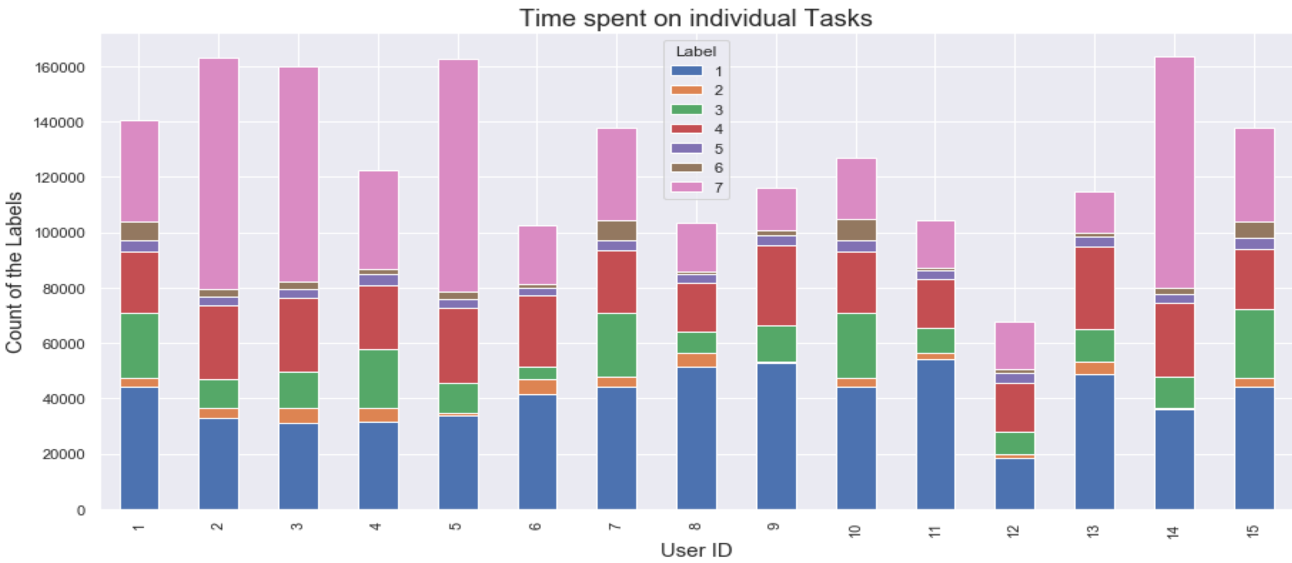
4.2. Results from relationship exploration between the columns

The table of hypothesis with the conclusions:

No	Hypothesis	Conclusion
1	Time spent by each individual on task 1 (Working on Computer) is the maximum. As the Task 1 is being seen 608667 times in the data, and it is the highest.	Users 2,3,5,14 have spent more time on Task 7 than on Task 1, the stated Hypothesis 1 is false.
2	Time spent on Task 6 (Walking and Talking with Someone) by the individuals is minimum. Tasks 2,5,6 have the approximately a smaller number of observations on the whole data but task 6 has the least.	Task 6 is not the least performed action when it viewed individually. Hence, the hypothesis 2 is False
3	Relationship between the acceleration in the x-axis and z-axis is almost null. As the correlation is found to be 0.009827 between the two columns.	No relationship can be plotted between the acceleration in x-axis and z-axis, Hence the hypothesis 3 is True.
4	Relationship between the acceleration in the x-axis and y-axis is minimum. As the correlation is found to be 0.363657 between the two columns.	A positive mild upward linear relationship is detected between the two columns. Hence the hypothesis 4 is True
5	Relationship between the acceleration in the y-axis and z-axis is minimum and is similar to the relationship between the acceleration in the x-axis and y-axis. As the correlation is found to be 0.345655 between the two columns which is similar to the correlation between the x_acceleration and y_acceleration column.	A mild positive relationship is found between the two columns, but there are more outliers. Hence the hypothesis 5 is True.
6	The maximum value of acceleration is for Task 1 in X-axis. Because, the Maximum value of acceleration in the X-axis is found to 3828 and Task 1 has most values (608667).	From the graph the hypothesis is True, as the maximum and minimum values are for Task 1
7	The values of acceleration in the y-axis is plotted against the Labels, to determine the dependency.	The Tasks other than the first one has the values of maximum and

		minimum values in the same range, The Task1 has extreme values.
8	The values of acceleration in the z-axis is plotted against the Labels, to determine the dependency.	The maximum and minimum values for all tasks have different values and task 1 has the extreme value in both maximum and minimum values.
9	The labelling of the tasks is more dependent on the x_acceleration than on the y and z axis acceleration. As, the correlation of x-axis with Label is the highest.	The labels are dependent on all x, y, z axis acceleration. Hence, the hypothesis 9 is False
10	Time spent on Task 6 on a whole is the minimum. As the individual evaluation of the graph predicted that the Task 6 is least performed.	After grouping the activities for the individuals, it can be seen that Task 5 is the least performed Task. Hence hypothesis 10 is False.

1. From Hypothesis 1, it is found that the activity of working on computer is the maximum.



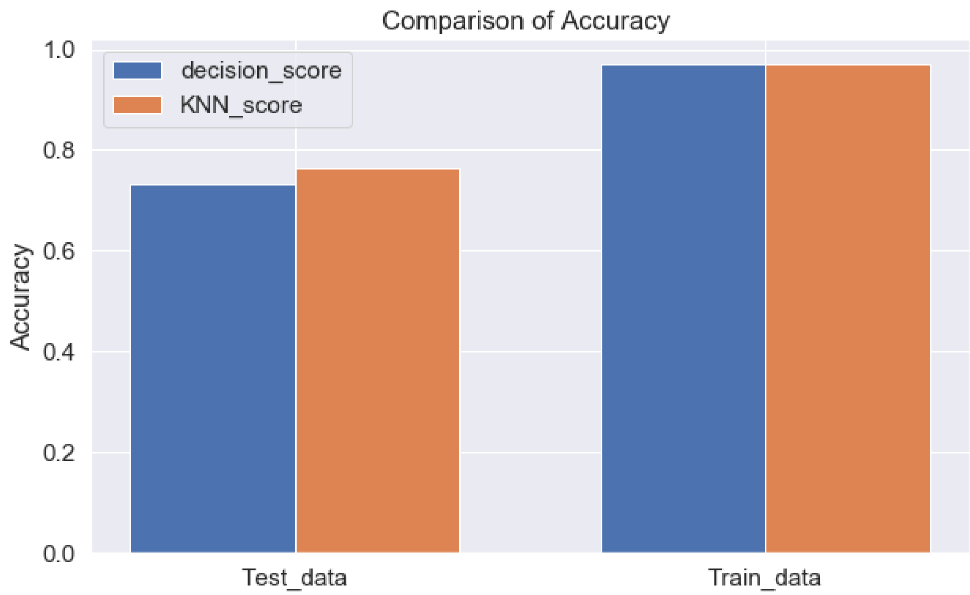
2. From hypothesis 2, it found that task 2 and task 6 are least performed by alternative users.

4.3. Results from Data modelling

Decision tree: The accuracy of the model is 73.15% and remains the same after applying the hill climbing method as there are 4 features selected and plotted against the Task feature.

KNN Classifier: Elbow method will determine the optimal value of K (Number of clusters) i.e. 8 for the accelerometer dataset. The accuracy increased by 0.4% after parameter tuning effect, hence the accuracy on the data is 76.45%

Comparison of Decision tree and KNN Classifier:



1. Accuracy of Decision tree is 73.15% and KNN Classifier is 76.45% on the Test data, and the accuracy is same for the train data.
2. Accuracy from the KNN Classifier is slightly higher than that of Decision tree.
3. The computational complexity for KNN is comparatively more for the Decision tree on the dataset.
4. The precision score for Decision tree is 0.52 and KNN classifier is 0.59.

Recommendation: K-Nearest Neighbours is a better classifier on the Accelerometer dataset.

5. Discussion

The sensor should be customized to perform the experiments without having the users to note down the order of the activities and do the start and stop of the activities. The device should automatically start collecting the data when the activity begins. If the activities apart from the labelled once should be collected, newly detected activities should be backtracked and given a new Label. The activities in the label are not quite distinguishable but can be better interpreted with the help of either the audio or video data along with the acceleration.

6. Conclusion

Most of the time of Individual users is spent on the Computer and next most of the time is spent on Talking while Standing, which are not physical exercises. Hence causing harmful effects in a long term. The sensor can be used in rehabilitation centres to keep a constant check on the patients. The device is currently being used for the research purposes, but it can be used for the clinical purposes. When the device is operated independent of the Users start and stop button. Accelerometer can be used for medical applications such as, monitoring people based on the gait parameters (Culhane, O'Connor, Lyons & Lyons 2005). Parkinson's disease is one of the research questions opened up, where the drug treatments are varied to assess the mobility and twitching of the body parts in the patients.

7. References

Brownlee, J 2018,'A Gentle Introduction to a Standard Human Activity Recognition Problem', *Machine Learning Mastery,* blog post, 12 September, viewed 11 June 2020, <https://machinelearningmastery.com/how-to-load-and-explore-a-standard-human-activity-recognition-problem/>

Casale, P Radeva, P & Pujol, O 2011, 'Human Activity Recognition from Accelerometer Data Using a Wearable Device' *Pattern Recognition and Image Analysis: 5th Iberian Conference, IbPRIA 2011, Las Palmas de Gran Canaria, Spain, June 8-10, 2011. Proceedings*, pp.289-296

Culhane, KM O'Connor, M Lyons, D & Lyons, GM 2005, 'Accelerometers in rehabilitation medicine for older adults', *Age and Ageing*, vol. 34, pp. 556–560

Images

Brownlee, J 2018, *A Gentle Introduction to a Standard Human Activity Recognition Problem*, photograph, accessed 10 June 2020, <https://machinelearningmastery.com/how-to-load-and-explore-a-standard-human-activity-recognition-problem/>

Tadi, MJ Koivisto, T Pänkäälä, M & Paasio, A 2014, International Journal of Biomedical Imaging, photograph, accessed 10 June 2020, https://www.hindawi.com/journals/ijbi/2014/690124/

Dataset

Dua, D. and Graff, C. (2019). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.