Thesis Dataset Report

Data Analysis for Physical Activity Monitoring

Rameesha Asghar Khan
University of Stavanger, Norway
r.asghar@stud.uis.no

ABSTRACT

This report is about the dataset used in the thesis project 'Data Analysis for Physical Activity Monitoring'. Physical activity monitoring is an emerging research area in wearable computing. In this report we will delve into the open source dataset PAMAP2; its layout, fields and collection methods.

KEYWORDS

Exploring dataset, Physical activity monitoring, PAMAP2

1 INTRODUCTION

Wearable computers and gadgets are making it possible for a common person to monitor his own physical activity and thus health, which was not possible in recent past. Monitoring physical activity is an emerging shiny field attracting new research. Our thesis is about analyzing the data we get from these wearables. The dataset we are exploring is PAMAP2 Physical Activity Monitoring Dataset. It is an open source dataset made available at UCI repository [1].

This dataset is chosen because of extensive physical activities: both everyday household and sports, performed by 9 subjects wearing 3 IMUs(Inertial Measurement Units) and a heart rate monitor.

2 DATA COLLECTION

The data is collected from 9 test subjects, 8 male and 1 female. All of them were aged 27.22 +/- 3.31 years with BMI ranging 25.11 +/- 2.62 kgm^{-2} [2].

Three inertial measurement units (IMUs) and a heartrate monitor were worn by these subjects. Sampling frequency of the IMUs is 100Hz; i.e. data is collected at every 0.01 second. The sampling frequency of heart rate monitor is 9Hz.

The placement of the IMU sensors was: [1]

- 1 IMU over the wrist on the dominant arm
- 1 IMU on the chest
- 1 IMU on the dominant side's ankle

The participants had to perform 12 protocol and 6 optional activities. The list of activities is given in Table 1

Protocol Activities	Optional Activities
Lie	Watching TV
Sit	Computer Work
Stand	Car Driving
Iron	Folding laundry
Vacuum clean	House cleaning
Descend stairs	Playing soccer
Ascend stairs	
Normal walk	
Nordic walk	
Cycle	
Run	
Rope jump	

Table 1: Activities performed by the test subjects

These activities and their mapping is given in the README file

3 DATA FORMAT

The dataset contains two sub-folders: Protocol and Optional. In addition to that following five pdf files are included:

- readme
- DataCollectionProtocol
- DescriptionOfActivities
- PerformedActivitiesSummary
- subjectInformation

The titles of the files are self-explanatory.

The protocol folder contains information from all 9 subjects, whereas, the optional folder contains data from only five test subjects.

All files contains data in 54 columns. The distribution of information in columns, as described in Readme file, is as below [1]:

- 1 timestamp (s)
- 2 activityID
- 3 heart rate (bpm)
- 4-20 IMU hand
- 21-37 IMU chest
- 38-54 IMU ankle

Further explanation on the IMU columns is given in the attached files.

Protocol folder. contains 2872533 entries, total combined data of all subjects. We know that we should ignore the transitional entries, with activityID = 0, so after that we are left with 1942872 rows.

Number of data entries per test subject can be seen in the Figure 1

Supervised by Tomasz Wiktorski.

Master Thesis in Computer Science, UiS,
2019.

Figure 1: Number of protocol activities by subjects

Optional folder. contains 977972 rows and 54 columns, total combined data of all subjects. After removing transitional data, we are left with 782081 entries. Figure 4 shows number of data entries per test subject.

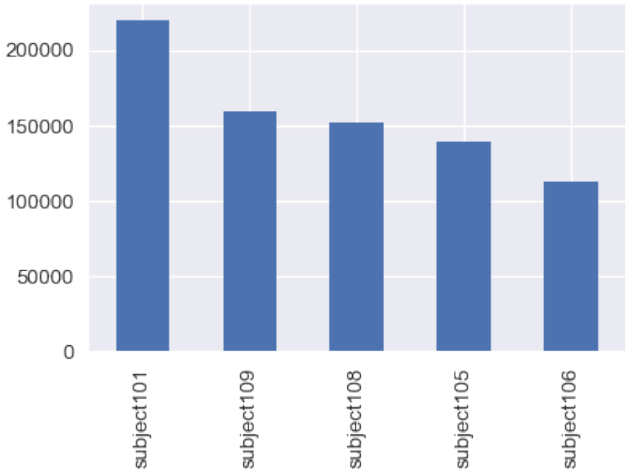

Figure 2: Number of optional activities by subjects

4 DATA SUMMARY

Using plots we can visualize different features of the dataset in order to understand better.

Performed Activities. Using crosstab function of the Pandas we have created the summary of the activities performed by the test subjects.

activityID	1	2	3	4	5	6	7	9	10	11	12	13	16	17	18	19	20	24	All
subject																			
subject101	27187	23480	21717	22253	21265	23575	20265	83646	0	54519	15890	14899	22941	23573	27114	54089	0	12912	469325
subject102	23430	22345	25576	32533	9238	25108	29739	0	0	0	17342	15213	20683	28880	0	0	0	13262	263349
subject103	22044	28761	20533	29036	0	0	0	0	0	0	10389	15275	20325	27975	0	0	0	0	174338
subject104	23047	25492	24706	31932	1	22699	27533	0	0	0	16694	14285	20037	24995	0	0	0	0	231421
subject105	23699	26864	22132	32033	24646	24577	26271	0	110883	0	14281	12727	24445	33034	0	28488	0	7733	411813
subject106	23340	23041	24356	25721	22825	20486	26686	0	61777	0	13291	11272	21078	37744	21786	28714	0	256	362373
subject107	25611	12282	25751	33720	3692	22680	28725	0	0	0	17646	11618	21552	29499	0	0	0	0	232776
subject108	24165	22923	25160	31533	16532	25475	28888	0	68725	0	11683	9655	24292	32990	23650	41691	18126	8806	414294
subject109	0	0	0	0	0	0	0	0	68550	0	0	0	0	0	27328	34206	28789	6391	165264
All	192523	185188	189931	238761	98199	164600	188107	83646	309935	54519	117216	104944	175353	238690	99878	187188	46915	49360	2724953

Figure 3: Performed activities summary

Figure ?? and 5 show the total number of times the activities performed by the test subjects

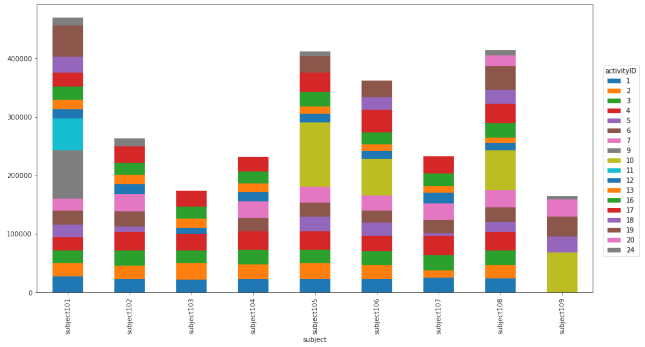

Figure 4: Performed activities summary stack graph

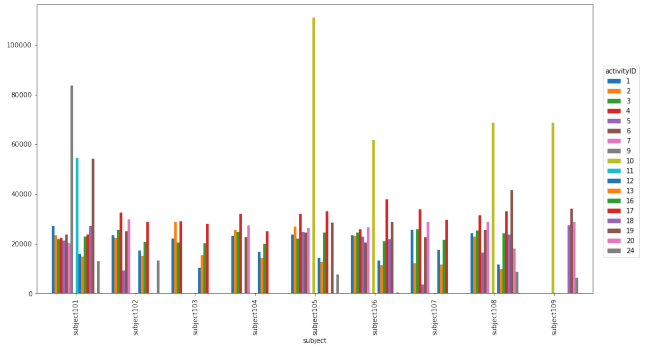

Figure 5: Performed activities summary bar plot

REFERENCES

[1] Attila Reiss. 2012. PAMAP2 Physical Activity Monitoring Data Set. http://archive.ics.uci.edu/ml/datasets/pamap2+physical+activity+monitoring

[2] Attila Reiss and Didier Stricker. 2012. Introducing a New Benchmarked Dataset for Activity Monitoring. (2012). https://docs.google.com/viewer?a=v&pid=sites&srcid=ZGVmYXVsdGRvbWFpbnxhdHRpbGFyZWlzc3xneDozNjg1NDY1NGEzOGM1ZWI0