First Iteration

The ActivPal accelerometer was used to track acceleration of subjects during the lab session. The ActivPal accelerometer recorded data in resample rate of 20 Hz. Each activity duration was 5 minutes equals to around 6000 rows in data. Each row contains X-axis, Y-axis, Z-axis values with datetime indexes. The X, Y, Z-axis was processed to gravitational acceleration with following formula {PLACEHOlder}. Accurate assessment of activity duration requires analysis in small segments, but segments of larger size might carry more meaningful information on the type of activity and improve the recognition of Phyiscal activity(PA). (link naar paper). Multiple Random Forest models were created to track the best segment size. The data going into these models were segmented from 1 to 14 seconds with steps of 0.1 seconds. Accuracy, recall and precision were considered on deciding which 3 segment sizes to uphold. In the lab sessions the lab assistant logged each activity start and end time of every respondent which was used to label the time segments with the correct activity. Within each time segment the features for the model was created. These features consist of the mean and standard deviation of the X, Y and Z axis.

The splitting of the train, validation and test sets was executed in two steps. First 3 of 25 subjects was extracted from the subjects to be used as test dataset. These 3 were most representative against the training dataset which was determined with the algorithm described in chapter {PLACEHOLDER}. Then the data of remaining 22 subjects was split in to training (80%) and validation (20%) dataset which was done with training_test_split function from Sklearn.

The validation dataset was used in development of Random Forest model to determine which configuration gave the best results. The results which were used were from accuracy, f1, recall and precision from the prediction on the validation dataset. K-fold cross-validation was also used to determine how the model reacted to unseen data in 5 folds. The data which was used was combination of training and validation dataset.

The Random forest model was developed with the processed data previously described as input. The activities the respondents performed in the lab session are as follows: cycling light, cycling heavy, walking, running, jumping and walking on stairs. Jumping and walking on stairs were excluded because of missing Vyntus data, therefore missing ground truth of MET-values. The number of trees in the random forest model was decided upon experimentally. A range of 10 to 200 random forest models were created with each having different number of trees. The number of trees was decided upon by the best accuracy score.

Met opmerkingen [W(1]: Is een afkorting, maar de hele betekenis staat nog niet in de tekst (volgens mij)

Met opmerkingen [S(2]: dit zegt dat grotere segments recognition of PA kunnen verbeteren maar in de volgende zin staat er "therefore data was cut in time segments to classify activity in a smaller time range"

Met opmerkingen [T(3]: Hele sterke introductie

Met opmerkingen [W(4]: Is een afkorting, maar de hele betekenis staat nog niet in de tekst (volgens mij)

Met opmerkingen [S(5]: dit zegt dat grotere segments recognition of PA kunnen verbeteren maar in de volgende zin staat er "therefore data was cut in time segments to classify activity in a smaller time range"

Met opmerkingen [S(6]: ik zou deze zin weglaten omdat het een contradictie is en niet uitlegd hoe de time segment is bepaald

Met opmerkingen [W(7]: Is vanaf `accuracy` een begin van een nieuwe zin?

Met opmerkingen [S(8]: is er 1 segment size gekozen of meerdere sizes?

Met opmerkingen [T(9]: Hele sterke introductie

Met opmerkingen [S(10]: Is populatie de juiste term of steekproefpopulatie?

Met opmerkingen [W(11]: Leg dit stuk uit hoe we dit hebben gedaan voor activity recognition EN met prediction. Dus ook dat we voor beide modellen verschillende representatieve test users hebben gekregen na het toepassen van het algoritme.

Met opmerkingen [A(12R11]: Gaan we dit nog in een aparte hoofdstuk bespreken of niet?

Met opmerkingen [T(13]: Goed onthouden dat dit vooraf wel moet zijn omschreven wat vyntus data is.

Met opmerkingen [A(14R13]: Dit moet toch in subject design of intrudoction to subjects worden beschreven?