Fourth iteration

The activPAL accelerometer was used to track acceleration of subjects during the lab session. The activPAL accelerometer recorded data in resample rate of 0.05S (Why activPAL?, z.d.). Each activity duration was 5 minutes equals to around 6000 rows in data. Each row contains X, Y and Z-axis values with datetime indexes. The X, Y and Z-axis were processed to gravitational acceleration with equation 1.

$$G = \frac{a - 127}{63} \qquad \text{[equation 1]}$$

Where G is gravitational acceleration and a is scaled acceleration from ActivPal Acceleration of X, Y or Z axis. 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 Physical activity (PA) (Bonomi, 2009). Multiple random forest (RF) (Tin Kam Ho, 1995) classifiers 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 from k-fold cross validation 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 dataset was split in train, validation and test following the steps which are described in chapter 'Train, Validation and Test dataset'. K-fold cross validation was applied on the validation dataset to determine which segment sizes gave the best result. At the same time the number of trees was decided upon experimentally. A range of 10 to 200 RF classifiers were created with each having different number of trees. The best number of trees was decided upon by looking at the accuracy score. The best segment sizes were decided upon looking at metrics accuracy, precision and recall on the validation dataset and the K-fold cross-validation. K-fold cross-validation was to determine how the model reacted to unseen data in 5 folds. The data used with k-fold cross-validation exist of combination of training and validation dataset

The RF classifiers 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 ability of the model to distinct activities from each other were analysed visually with confusion matrix on the validation dataset.

Bonomi, A. G., Goris, A. H., Yin, B., & Westerterp, K. R. (2009). Detection of type, duration, and intensity of physical activity using an accelerometer. *Medicine & Science in Sports & Exercise*, *41*(9), 1770-1777.

Why activPAL? (z.d.). PALT. http://www.palt.com/why-activpal/

Tin Kam Ho, "Random decision forests," Proceedings of 3rd International Conference on Document Analysis and Recognition, Montreal, Quebec, Canada, 1995, pp. 278-282 vol.1, doi: 10.1109/ICDAR.1995.598994.