# Human Activity Recognition based on the accelerometer data

Mikhail Kilianovski

October, 2020

## Overview

The task was to build a classification model for physical activities. Baseline shows that 3 out of 4 activities are linearly separable in simple feature space. Problems arouse with 'stairs' and 'walking' classes. This aligns with the intuition - these two activities are quite similar. To distinguish these two classes, more features were extracted from raw time series. Logistic Regression and Random forest models were trained on these feature vectors. Uninformative features were filtered during feature selection process for each model. As a result, models that confidently distinguish between 4 proposed activities were built. Concerns regarding generalization ability and ideas for further work are described in the conclusion section.

## Validation setup

I use 5-fold stratified cross validation. Stratification is needed to preserve class proportion in each fold. 3 performance metrics are calculated on each fold and averaged:

- **Accuracy**. This metric could be misleading due to class imbalance, but it is intuitively interpretable and all related papers reports it.
- **F1 macro**. F1 metric is a good choice when classes are imbalanced. In Multi-Class Classification problem, F1 macro is calculated as a mean of F1 scores for each class.
- **F1 min**. Calculated as a minimum value of F1 scores for each class. Motivation for this is the following: if we want to cover each class, such metric will quantify performance on the worst case.

## Baseline

3 features were generated for the baseline model - means for every axis(x,y,z). This provided a strong baseline - 'idle', 'running' and ('walking' + 'stairs') categories are separated almost perfectly. However, it is hard to separate 'walking' from 'stairs' instances.

# Feature engineering

Feature engineering is based on related work on human activity recognition [1, 2]. First, magnitude of x,y and z is calculated for each time step and later used as 4th timeseries $m$. Then, for each axis the following features are calculated:

- Mean
- Standard deviation
- Min
- Minmax (max minus min)

Pearson correlation coefficient between (x,y), (y,z), (x,z) axis is calculated. In total, there are 19 features.

# Numerical results

| Classifier | Features | Accuracy($\sigma$) | F1 macro($\sigma$) | F1 min($\sigma$) |
|---|---|---|---|---|
| Logistic regr. | 3 means (x,y,z) | .978(.003) | .850(.015) | .442(.055) |
| | all_features | .991(.003) | .948(.018) | .809(.067) |
| | all_features_poly2 | .997(.001) | .986(.008) | **.947(.030)** |
| | poly2_top23 | .996(.002) | .978(.011) | .919(.040) |
| Random Forest | 3 means (x,y,z) | .987(.001) | .912(.005) | .672(.017) |
| | all_features | .999(.001) | .994(.006) | **.978(.021)** |
| | top7_features | .997(.002) | .989(.010) | .959(.036) |
| Random Forest (OOB preds ) | all_features | .998(-) | .990(-) | .965(-) |

Table 1: Classification quality comparison

## Logistic Regression

Logistic regression with L2 regularization on all features provides good results. Standard scaling were used to improve convergence. Adding polynomial interactions of degree 2 improves quality significantly. L1 regularization were used to select the most informative features.

## Random Forest

Random Forest works better on this problem in all subset of features. Selecting top 7 features from 19 does not decrease performance significantly. Also, out-of-bag predictions were calculated and metrics does not deviate significantly from results on cross-validation

## Conclusion and further work

We can conclude that proposed 4 activities are easily separable even in simple feature spaces. However, if models will be used on data from different domain (e.g. for other person, on different smartphone model etc.) these results could not be reliable.

One idea that could improve quality on a data with more variability is using smoothing techniques for signal denoising as proposed in [2]. This includes median filters and Butterworth filters.

Other direction of improvement might be working with peaks, as showed in [1]. This includes detection of peaks and building features from this information.

More recent works propose applying neural networks, especially CNNs. If we have a lot of data, these approaches may be more effective that hand-crafted features. Moreover, neural networks open ways to utilize unlabeled data in unsupervised/semi-supervised fashion.

Also, I conducted experiments with applying SVM with RBF kernel on raw data with smaller window size. Idea is the following: creation of smaller windows dramatically increases number of training observations and reduces dimension of feature space (which is raw readings). If we use window size of 15, we increase number of training observations in 16 times. However, this approach is computationally heavier and requires further analysis.

## References


[1] Akram Bayat et al. *A Study on Human Activity Recognition Using Accelerometer Data from Smartphones*. 2014. URL: `https://www.sciencedirect.com/science/article/pii/S1877050914008643`.

[2] Dario Ortega-Anderez et al. "A multi-level refinement approach towards the classification of quotidian activities using accelerometer data". In: *Journal of Ambient Intelligence and Humanized Computing* 10.11 (Nov. 2019), pp. 4319–4330. ISSN: 1868-5145. DOI: `10.1007/s12652-018-1110-y`. URL: `https://doi.org/10.1007/s12652-018-1110-y`.