Machine Learning Project 2014-2015 (Oct)

Who has my phone?

1 Introduction

Every person walks in a unique manner. This uniqueness is what we want to quantify in this project. Nowadays, all smart phones have a range of sensors build-in such as accelerometers, gyroscopes, or GPS. We are going to build on the data generated by these sensors to solve our task at hand: Can you detect which person is walking around with a phone based on how the phone behaves in their trouser's pocket? For this to be possible, it is necessary to apply machine learning techniques to detect a combination of characteristics that is unique to each person.

2 Problem and Approach

Your task is to classify 10 unlabeled walks based on a training set of 58 walks. In total 30 persons have participated in the data collection, but in this project you only have to consider three labels: Wannes, Leander and Other.

Data is collected using an Android app sampling the accelerometer, gyroscope and magnetometer at a speed of $\pm$50Hz (the .json-files). Acceleration is measured along three axes according to the phone coordinate system with the gravitational component removed. Since this would mean that the data are different depending on how you put the phone in your pocket (e.g., screen in- or outside), this is not an ideal representation. Therefore, the data is preprocessed to express acceleration according to the geographical coordinate system (the .csv-files):

X	Acceleration along the West-East axis
Y	Acceleration along the (magnetic) South-North axis
Z	Acceleration orthogonal to the earth surface

In order to use raw accelerometer data for classification techniques like Naive Bayes, SVMs or Decision trees it is often necessary to first extract features from the data (see [1, 2, 3, 4] for

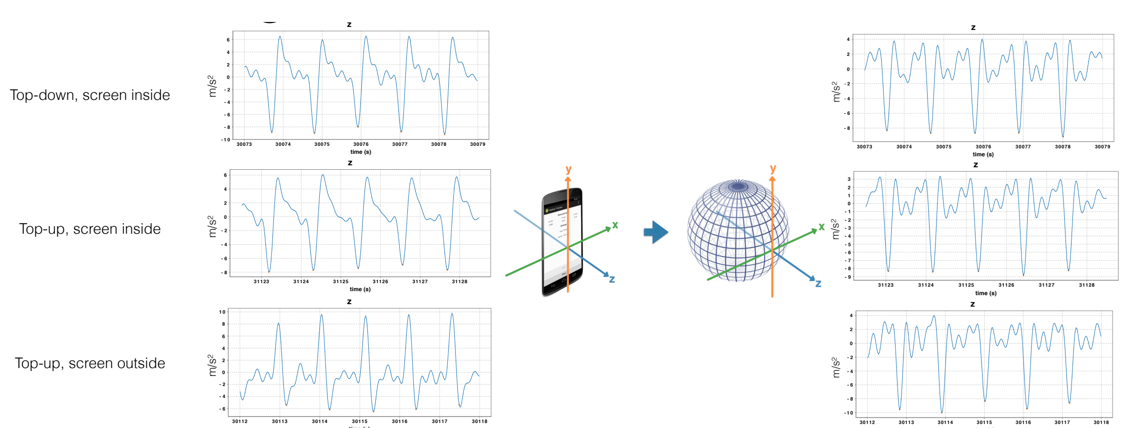

Figure 1: Transformation from phone coordinate system to geographical coordinate system.

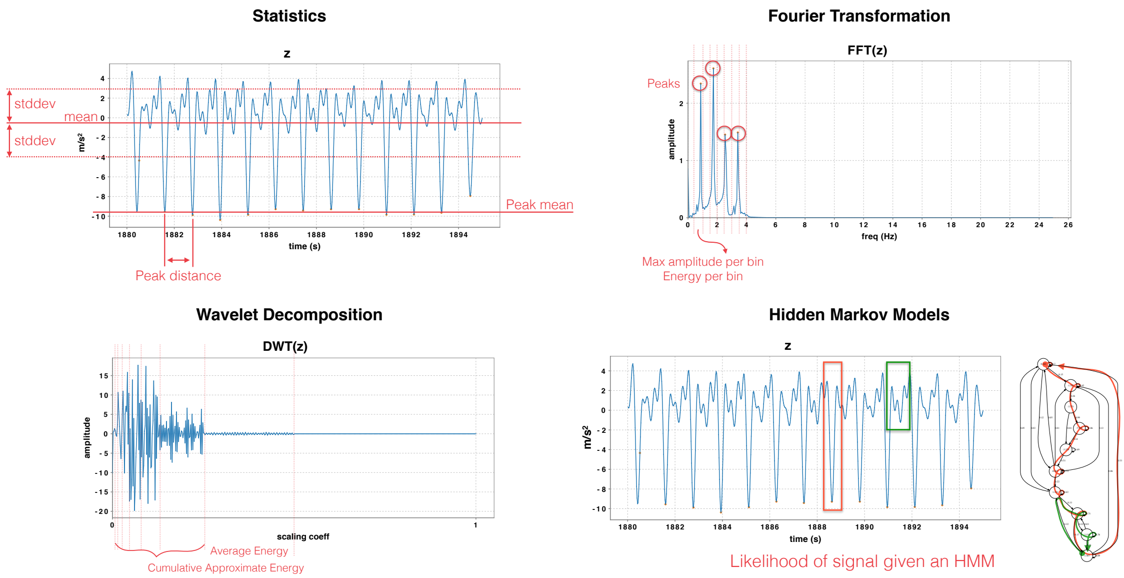

Figure 2: Example features you could consider.

inspiration). Useful features are for example simple statistics like average, variance or kurtosis, average peak distance or height, peaks in the frequency domain, energy in a wavelet component, HMM likelihood, ...

A file is not guaranteed to be pure walking data. Putting the phone in a pocket, turning in the hallway or taking the phone out of a pocket are also part of the data. Therefore, it will be necessary to device a method to extract windows of clean walking data. To achieve this you can, for example, train a separate classifier to recognize walking in general and use this classifier for pre-processing the data.

Additionally, you also get a test data set where the walks are not labeled. Your task is to predict the labels for these walks. Extra points will be rewarded to the teams with the highest scores (divided by the number of teams that achieve this score).

3 Tasks

3.1 Form Groups — By October 24th

You can work on this project in groups of two and divide the work. Form the groups as quickly as possible and email the names of the members of your group to davide.nitti@cs.kuleuven.be. You work on this assignment alone or in team, **do not share your experimental results or implementations with your fellow students**.

3.2 Literature Study

Familiarize yourself with the concept of machine learning for accelerometer data and activity recognition. Get a feeling for what has been done before in this context, and what are the open questions. Divide the work between the two people in the team.

3.3 First Report **By November 7th**

In a first phase of the project you explain in a brief report (about 2 pages) how you will address these tasks. Describe what literature you have read and what you have learned from it. Give examples of questions that you want to find answers for, and give an idea of how you intend to find the answers (What data representation, features and machine learning techniques will you use? How will you evaluate your classifier?). Please send your report in PDF format to the aforementioned email address. The quality of this report influences your final score for the project, so do your best to come up with a good plan. After the reports are handed in, you will get feedback on your report, and, where necessary, we may give more concrete guidelines on how to proceed.

3.4 Machine Learning

Use at least two different machine learning techniques to classify the data. Make sure that you use a correct experimental setup to assess the accuracy of the classifier (e.g., cross-validation, test/train split).

3.5 Analyze results

What is the accuracy you can achieve when applying cross-validation on the training data? Do you understand why your parameters work? What is the computational cost of pre-processing, learning and classifying? What is your prediction for the 10 unlabeled walks?

3.6 Final Report and Prototype **By December 12th**

Report: Write a report with your findings.
Mail your final report, describing details of your approach and the results obtained:

- Clearly state the **questions** you address with your research
- Describe how you try to **answer** them (what experiments were performed, how do you measure success, etc.),
- Write out the **conclusions** you draw from your experiments together with a scientifically supported motivation for these conclusions. Such a motivation should include descriptions of used techniques and performed experiments and a report and discussion of the results of these experiments. Be careful to report concrete numbers and custom formulas (avoid conclusions using *weasel words*).
- Report your prediction for the unlabeled walks.
- Report the total **time you spent** on the project, and how it was divided over the different tasks mentioned.

The total length of your report should be at most 10 pages.

Prototype: Include the code you used to perform the machine learning and experiments with the final report. We will test unseen examples using your prototype to assess the accuracy. You are free to use any mainstream programming language and any publicly available toolbox. Make sure that:

- The code is intuitive to run and includes a short README file with the necessary steps.

- The default setting is to use the classifier you have found to be the best.
- Your code is self-contained and includes all dependencies.
- Your code is print-friendly (e.g., max 80 columns).
- It is easy to give as input a path to a new file and let your software print the classification (this will be used to test your software on an unseen walk).

3.7 Peer assessment — By December 12th, individually

Send by email a peer-assessment of your partner's efforts. This should be done on a scale from 0-4 where 0 means "I did all the work", 2 means "I and my partner did about the same effort", and 4 means "My partner did all the work". Add a short motivation to clarify your score. This information is used only by the professor and his assistants and is not communicated further.

3.8 Discussion — By December 19-20

There will be an oral discussion of your project report in week 13 where you will also get the chance to demo your prototype.

Time

The time allocated for completing this project is **60 hours** per person. This *does not* include studying the Machine Learning course material, but it *does* include the literature study specific for this project's topic.

Questions

Please direct any questions that you may have about the project to the Toledo forum.

Good luck!

References

[1] Mark V. Albert, Konrad Kording, Megan Herrmann, and Arun Jayaraman. "Fall Classification by Machine Learning Using Mobile Phones". In: *PLoS ONE* 7.5 (2012).

[2] Jonathan Lester, Tanzeem Choudhury, Nicky Kern, Gaetano Borriello, and Blake Hannaford. "A Hybrid Discriminative/Generative Approach for Modeling Human Activities". In: *Proceedings of IJCAI*. 2005.

[3] Stephen J Preece, John Yannis Goulermas, Laurence PJ Kenney, and David Howard. "A comparison of feature extraction methods for the classification of dynamic activities from accelerometer data". In: *Biomedical Engineering, IEEE Transactions on* 56.3 (2009), pp. 871–879.

[4] Nishkam Ravi, Nikhil Dandekar, Preetham Mysore, and Michael L Littman. "Activity recognition from accelerometer data". In: *Proceedings of AAAI*. 2005.