KU Leuven
Departement of Computer Science

Machine Learning

Who has my phone?

Group:

Thijs Dieltjens
Bram Geelen

Academiejaar 2014 – 2015

1 Introduction

For the given assignment we need to be able to detect which person is walking around with a phone. This detection will be based on previous, labeled data, generated by other phones. To do this, we will use different Machine Learning techniques, which we researched in a concise literature study. We also present a general approach to the implementation of the problem.

2 Literature study

The most useful paper we found in our literature study was [1]. This paper reported algorithms to detect a big range of physical activities from data, Eg. walking, tooth brushing…It also discusses feature computation, which is heavily referred to in other literature. The writers applied four classifiers, and found that the best classifier were decision trees, more specifically the C4.5 algorithm.

In [2], the goal of the research was similar: the detection of different falls. However the tracked action is different from other papers, the used algorithms stay the same. The results of this paper are very promising, but this may lie in the fact that a fall is a more abrupt action, whereas running is a much *smoother* action. Five classification techniques are used. In this case Support Vector Machines (SVM) and Sparse Multinomial Logistic Regression (SMLR) were the best classifiers. The paper also lists some extra used features.

[3] describes the usage of features and expands on how to select the best features and drop the less useful ones, by *boosting* and *bagging*. In the paper, 600 features are reduced to 50, without losing 1% of accuracy. This reduction might deliver a big performance gain, although it means a big additional implementation. On top of this feature selection, it uses Hidden Markov Models (HMMs).

The paper [4] presents a comparison of 14 methods to extract classification features from accelerometer signals. The combination of time and frequency features is found to give the best results. Features derived from Fast Fourier Transform (FFT) outperformed the results derived from wavelet coefficients.

A nice coincidence is the same data frequency of 50Hz that appears in [5]. The research goal of this paper is mostly the same as in other papers, but the writers also applied several methods to have a new exhaustive set of meta-level classifiers. These contain Boosting, Bagging, Plurality Voting, Stacking with Ordinary-Decision trees (ODTs) and Stacking with Meta-Decision trees (MDTs). Plurality voting turned out to yield the best results.

The goal in [6] is exactly the same as the goal of the project. This paper is the only paper which mentions a distinction between left and right steps, which might also be useful data (especially when distinguishing between different walkers). The main features that are used in the research are correlation, histogram statistics and FFT coefficients. However, it is not specified which classifiers are used to make a prediction. Since the goal of the research is the same as ours, the performance of this paper is also relevant; the writers achieve an accuracy of at least 72 percent.

In addition to these aforementioned papers, the following papers turned out to be either less applicable, or contained less additional information although they were on-topic: [7], [8], [9], [10] [11].

3 Problem

The goal is to classify 10 unlabeled walks, based on a training set. The *walks* consist of data generated by accelerometers in mobile phones. To use the data, we need to be able to extract useful features.

Initially, we need to implement a method which extracts windows of clean walking data. To do this, we first need to look at how big the window size has to be. Perhaps it is possible to create windows for each step (as proposed by [6]). Afterwards, the data is partitioned in multiple windows. A classifier classifies the windows as *clean* or *unclean*. To do this, we will implement the classifier as a simple decision tree, k-nearest-neighbour or another base-level classifier (depending on which performs best).

With this clean data, we can then apply the algorithms used in the different papers. Most papers indicate that the following features can be useful;

- Time-domain features: Mean and standard deviation, median and other percentiles
- Principal frequency
- Spectral energy
- Mean dc, energy, entropy and correlation between different axes
- Magnitude of components of FFT

If these features do not suffice to make an accurate prediction, it is possible to expand this set with the histogram values described in [6].

On these selected features, we can apply classifiers. We will use at least decision trees and SVMs. If more classifiers are used, plurality voting might be implemented if useful.

Finally n-fold validation will be used to test our performance on the labeled walks.

The entire implementation will be implemented with the WEKA-toolkit in Java.

References

[1] L. Bao and S. S. Intille, “Activity recognition from user-annotated acceleration data,” in *Pervasive computing*, pp. 1–17, Springer, 2004.

[2] M. V. Albert, K. Kording, M. Herrmann, and A. Jayaraman, “Fall classification by machine learning using mobile phones,” *PloS one*, vol. 7, no. 5, p. e36556, 2012.

[3] J. Lester, T. Choudhury, N. Kern, G. Borriello, and B. Hannaford, “A hybrid discriminative/generative approach for modeling human activities.,” in *IJCAI*, vol. 5, pp. 766–772, 2005.

[4] S. J. Preece, J. Y. Goulermas, L. P. Kenney, and D. Howard, “A comparison of feature extraction methods for the classification of dynamic activities from accelerometer data,” *Biomedical Engineering, IEEE Transactions on*, vol. 56, no. 3, pp. 871–879, 2009.

[5] N. Ravi, N. Dandekar, P. Mysore, and M. L. Littman, “Activity recognition from accelerometer data,” in *AAAI*, vol. 5, pp. 1541–1546, 2005.

[6] J. Mantyjarvi, M. Lindholm, E. Vildjiounaite, S.-M. Makela, and H. Ailisto, “Identifying users of portable devices from gait pattern with accelerometers,” in *Acoustics, Speech, and Signal Processing, 2005. Proceedings.(ICASSP’05). IEEE International Conference on*, vol. 2, pp. ii–973, IEEE, 2005.

[7] C. Strohrmann, M. Rossi, B. Arnrich, and G. Troster, “A data-driven approach to kinematic analysis in running using wearable technology,” in *Wearable and Implantable Body Sensor Networks (BSN), 2012 Ninth International Conference on*, pp. 118–123, IEEE, 2012.

[8] J. Mantyjarvi, J. Himberg, and T. Seppanen, “Recognizing human motion with multiple acceleration sensors,” in *Systems, Man, and Cybernetics, 2001 IEEE International Conference on*, vol. 2, pp. 747–752, IEEE, 2001.

[9] B. M. Eskofier, M. Kraus, J. T. Worobets, D. J. Stefanyshyn, and B. M. Nigg, “Pattern classification of kinematic and kinetic running data to distinguish gender, shod/barefoot and injury groups with feature ranking,” *Computer methods in biomechanics and biomedical engineering*, vol. 15, no. 5, pp. 467–474, 2012.

[10] J. Lester, T. Choudhury, and G. Borriello, “A practical approach to recognizing physical activities,” in *Pervasive Computing*, pp. 1–16, Springer, 2006.

[11] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, “Activity recognition using cell phone accelerometers,” *ACM SigKDD Explorations Newsletter*, vol. 12, no. 2, pp. 74–82, 2011.