Utrecht University

Doctoral Thesis

Human Acitivity Recognition Using Accelerometer Data

Author:
R.Q. Vlasveld

Supervisor:
Dr. James Smith

A thesis submitted in fulfilment of the requirements for the degree of Master of Science

in the

Research Group Name
Department or School Name

January 2013

Declaration of Authorship

I, R.Q. Vlasveld, declare that this thesis titled, 'Human Acitivity Recognition Using Accelerometer Data' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: ______________________________

Date: ______________________________

“Thanks to my solid academic training, today I can write hundreds of words on virtually any topic without possessing a shred of information, which is how I got a good job in journalism.”

Dave Barry

UNIVERSITY NAME (IN BLOCK CAPITALS)

Abstract

Faculty Name
Department or School Name

Master of Science

Human Acitivity Recognition Using Accelerometer Data

by R.Q. Vlasveld

The Thesis Abstract is written here (and usually kept to just this page). The page is kept centered vertically so can expand into the blank space above the title too...

Acknowledgements

The acknowledgements and the people to thank go here, don't forget to include your project advisor...

Contents

List of Figures

List of Tables

Abbreviations

LAH	**L**ist **A**bbreviations **H**ere

Physical Constants

Speed of Light $c = 2.997\ 924\ 58 \times 10^{8}\ \mathrm{ms}^{-\mathrm{S}}$ (exact)

Symbols

a	distance	m
P	power	W (Js^{-1})
ω	angular frequency	$rads^{-1}$

For/Dedicated to/To my…

Chapter 1

Introduction

1.1 Problem statement

*** Give context problems, means to reach (e.g. monitoring of patients, getting conclusions over health and activity).

1.2 Literature review

Look into earlier application mentioned in the literature to this kind of problems. Look for similarities in the problem and address where the techniques used fail or are not applicable.

1.3 Learning LaTeX

Chapter 2

Techniques

2.1 Signal pro-processing and sensor fusion for timed patterns

2.2 Temporal Segmentation

This section will give an introduction and in-depth analysis of temporal segmentation. test. nog een test. test

2.2.1 Aims of segmentation

When processing and analyzing time series of data, e.g. motion measurements, stock market fluctuations or natural language, first a low-level division between the discriminative parts of the stream must be made. One can view this as splitting the series into the *atoms*, which are the building blocks of the total stream. These building blocks will be the aggregation of non-overlapping, internally homogeneous segments [**?**]. This means that the data points inside a segment should have some resemblance relation to each other and their difference lies between some boundary. The process of segmenting can be viewed as a subproblem to context analysis of time series. Temporal segmentation is closely related to temporal clustering, although it is a stricter, and simpler, process. Whereby clustering only restricts the data points on their distance relation (as used in a Voronoi diagram), within a segment the data points must also be contiguous.

The task of segmentation can be performed in a manual matter, by cutting and labeling parts of the stream into coherent parts. This would require human (expert) knowledge and does not yield a clear cut because of ambiguity. With increasing storage abilities and easier motion capture systems, there is a desire for automated systems which perform the segmentation task unsupervised. Some algorithms used have the (often desired) side-effect of also clustering the segments, such that classes of segments can be discovered in the time series. These algorithms would not only be able to make a distinction between walking, sitting and walking, but would also recognize the reappearance of the walking activity.

2.2.2 Formal definition

Formally, temporal segmentation is dividing a time series s, which consists of N samples $\mathbf{x}(1), \mathbf{x}(2), \ldots, \mathbf{x}(N)$ from $\mathbf{R}^d$. Individual *segments* are referenced by $s(a, b)$, consisting of the consecutive samples $\mathbf{x}(a), \mathbf{x}(a+1), \ldots, \mathbf{x}(b)$, $a \leq b$. Let $s_1 = s(a, b)$ and $s_2 = s(b+1, c)$ be two segments, then their concatenation is $s_1 s_2 = s(a, c)$. A segmentation S of s consists of a sequence of k non-empty segments $s_1 s_2 \ldots s_k = s$. This notation has close resemblance to the notation as used in [**?**].

As stated, informally each segment should be internally homogeneous. This can formally be measured with an cost function F, indication the heterogeneity of a segment. The overall aim is to minimize the cost F. The cost of a segment is a function from the data points and the number of data points $n = b - a + 1$ and is expressed as

$$\text{cost}_F(s(a, b)) = F(\mathbf{x}; n | \mathbf{x} \in s(a, b)) \tag{2.1}$$

The cost of a *k-segmentation* S is the summation of the costs of the k segments:

$$\text{Cost}_F(s_1 s_2 \ldots s_k) = \sum_{i=1}^{k} \text{cost}_F(s_k) \tag{2.2}$$

With the objective of minimizing the cost function, the optimal k-segmentation $S_F^{opt}(s; k)$ is the segmentation with minimal $\text{Cost}_F(s_1 s_2 \ldots s_k)$ over all possible k-segmentations.

The cost function, to calculate the heterogeneity of a (set of) segment(s), can be any function. A simple and natural function would be the sum of variances of the segments. The overall cost function would then be

$$\text{Cost}_V = \frac{1}{N} \sum_{i=i}^{k} \sum_{j=c_{i-1}+1}^{c_i} \|\mathbf{x}(j) - \mu_i\|^2 \tag{2.3}$$

where μ_i is the mean vector of data points in segment s_i.

2.2.3 Application in research fields

[CHARACTERISTICS OF HUMAN MOTION] temporal variability, invariance over time, metrics over actions.

[COMPUTER VISION]

[GRAPHICS/VIDEO]

[DATA-MINING]

[MODEL BASED]

To analyze time series it is often preferred to divide the stream in segments of correlated data. After dividing, each segment represent a period in time in which the same activity is performed. Or, stated otherwise, it results in transitions moments between activities.

Many fields of research have been active in the unsupervised segmentation of data. Many authors rely on a form of Principal Component Analysis (PCA), as used a.o. in [**?**]. Often PCA is used to reduce the dimensionality of the data being processed [REFERENCE] by only using the top r dimensions to describe the data set. It is observed that data series of simple motions have a lower dimensionality then complexer motions. When a simple (repetitive) motion is about to end and fluently transforms in a new motion, there will be a window of time in which a high dimensionality will be present, due to the new motion. After this period of transition, the dimensionality will decrease, since only the new simple motion is present in the window of time. The first algorithm of [**?**] is based on this principle.

Given a set of data points, a lower dimensional hyperplane can by constructed to which the data points can be projected. This projection on a lower dimension introduces a error to the original position. When the error is fixed, less dimensions are needed for simple motions in which movements of body parts are highly correlated. For segments in which the data points are lesser correlated, e.g. because of transition state, a higher degree of dimensions of the hyperplane is needed to represent the data with equal error degree.

[SKIPPING] By analyzing the derivative of the error rate, the algorithm is capable of selecting transition points and thus segmenting the data. When a data point d_i is more than $k_\sigma = 3$ times the standard deviation from the average over the previous data points $d_j, j < i$, then a cut is assigned to that frame.

A second approach in [**?**] uses the probabilistic variant of PCA (PPCA) to model the data set as a Gaussian distribution instead of ignoring the frames which do not fit in the subspace. Over windows of frames the mean and variance are calculated. In a forward manner the Mahalanobis distance of a new window of frames is calculated, which represents the likelihood of the new window belonging to the same segment as the original widow. When the distance decreases, the likelihood increases which happens when the motions in the becomes more homogeneous. When a peak in the distance is reached, the new window of frames indicates a heterogeneous collection of motions in the window and thus a low likelihood of membership and a indication of a transition. In order to distinct activities and subactivities (which require a subset of motions is a distinct activity) the algorithm is also processed backward over the data series.

The third algorithm in [**?**] is based on the observation that data points (frames) tend to form clusters in the space. These clusters are represented by k Gaussian distributions for which each the Expectation-Maximization (EM) algorithm estimates the mean m_j, covariance matrix $\sum_j$ and prior π_j. With all the Gaussian distributions estimated, the data points are assigned to the cluster with the highest membership likelihood. When two consecutive frames x_i and x_{i+1} belong to different clusters, a transition of activities is recognized. Note that this algorithm succeeds in segmenting the data and also labels the similar simple activities.

A drawback in this system, and many others which implement a variant of the k-means algorithm, is that the number of clusters k need to be predetermined. To cope with this, often the algorithm is performed multiple times for different values of k. Using some criteria, e.g. the Bayesian Information Criterion [**?**] or the Davies-Bouldin Index which guides k-means clustering as used in [**?**].

2.2.4 Principal Component Analysis

2.2.5 Segmentation as clustering

In the previous section the discussed methods all relied on the stream of data points and tried to find cuts the discriminate between successive different type of activities. An other approach is to consider the tasks of segmentation as a type of clustering [**?**]. In clustering the objective is to assign labels, or classes, to all the data points indication a similar type of activity. A clustering $\mathcal{L}$ is thereby more informative then a segmentation but is also harder to produce.

A clustering $\mathcal{L}$ is generated from a sequence of elements $\mathbf{X}$ which is decomposed in m disjoint segments, each belonging to one of the k classes. A segment $\mathbf{Y}_i \hat{=} \mathbf{X}_{[s_i, s_{i+1})}$ is composed of frames from position s_i to s_{i+1}. A vector $g_{ci} = 1$ indicates class membership if $\mathbf{Y}_i$ belongs to class c, otherwise $g_{ci} = 0$.

When regarding segmentation of human motion as a task of clustering the difficulty is to model the temporal variability of actions and defining a robust metric between temporal actions. To overcome this, [**?**] introduces Aligned Cluster Analysis (ACA), by minimizing

$$J_{ACA}(\mathbf{G}, \mathbf{s}) = \sum_{c=i}^{K} \sum_{i=1}^{m} g_{ci}\, dist_c(\mathbf{X}_{[s_i, s_{i+1})}) \tag{2.4}$$

The characteristic of ACA is that is enables segments to span over different number of data points, whereas the standard kernel k-means algorithm results in equally sized segments. [!!! TRUE?!] The second difference is that the kernel used in $dist_c$ to measure the distance from a segment to the class which it is assigned to uses the Dynamic Time Alignment Kernel [REFERENCE?] to measure between time series.

2.3 Clustering

2.4 Temporal pattern recognition

2.4.1 Dynamic Time Warping

Used to measure similarity between time sequences. Exact matching is high-cost, so approximations such as Minimum Bounding Rectangles are used.

2.4.2 k-means clustering

[EXPLAIN] divide data set n into k clusters.

Among many unsupervised clustering techniques, k-means is successfully applied to large data sets. It is simple to implement and linear in time complexity so computationally attractive [**?**]. A drawback of the method is that the results of the algorithm greatly depends on the initial configuration (the data points which will act as centroids) and the number of cluster k must be determined beforehand.

Generally, the k-means methods will minimize the squared error for a clustering $\mathcal{L}$ criterion which is defined as the distance from the data points the centroid for each cluster in $\mathcal{K}$. This is expressed as optimizing to a local optimum the energy function

$$e^2(\mathcal{K}, \mathcal{L}) = \sum_{j=i}^{K} \sum_{i=1}^{n_j} \|\mathbf{x}_i^{(j)} - \mathbf{c}_j\|^2 \tag{2.5}$$

There are several limitation on the k-means method. One of these is that only spherical shapes of cluster can be generated. One of the extensions is kernel k-means [**?**], which implicitly projects the data points to a higher dimension and thereby is able to form irregular shaped cluster.

2.4.3 Self-organizing Map

2.4.4 Support Vector Machine

2.4.5 Naïve Bayes

2.5 Unsupervised clustering of temporal patterns

Chapter 3

Experiments

3.1

Chapter 4

Results

4.1

Chapter 5

Discussion

5.1

Chapter 6

Conclusions

6.1

Appendix A

Appendix Title Here

Write your Appendix content here.