

# Masters Thesis Lukas Kohlhasse

Lukas Kohlhasse

September 16, 2017

## Abstract

<sup>1</sup> I am an abstract that currently does not contain any content.

EdN:1

<sup>2</sup>

EdN:2

---

<sup>1</sup>EdNOTE: Make Abstract

<sup>2</sup>EdNOTE: change Title to something reasonable, still have to do good titlepage as well

# 1 Introduction

Okay we want some motivation stuff here.

Why do we want to segment data? (Useful for learning stuff, making datasets etc.) Why do we want to do so automatically etc. <sup>3</sup>

EdN:3

## 1.1 Related Work

Standard related work stuff here <sup>4</sup>

EdN:4

<sup>5</sup> SFA is nice, for reasons (universality), and it is also an unsupervised technique to get some classification data from a timeseries. Thus we want to try to use it for segmentation of timeseries data as well.

EdN:5

---

<sup>3</sup>EdNOTE: Make Introduction

<sup>4</sup>EdNOTE: Start Related Work here

<sup>5</sup>EdNOTE: Make Transition nicer

## 2 Basic Method

The basic Idea is that we first use SFA, then do clustering on the slow features, and then extract boundary points

### 2.1 Slow Feature Analysis

First you whiten the data, then you make temporal difference vectors, then do PCA take k least significant features, those are your feature vectors. Multiply with input data (centered I think) and you get your feature values

### 2.2 Spectral Clustering

Make Similarity Matrix, then do spectral Clustering on it. <sup>6</sup>

EdN:6

### 2.3 Find decision boundaries

Easy for case of 2 clusters, semi doable for 3 clusters, hard for n clusters. Will go into later

<sup>7</sup> Basic idea is clear, but the devil is in the details

EdN:7

---

<sup>6</sup>EDNOTE: Obviously explain spectral Clustering here

<sup>7</sup>EDNOTE: Transition

### 3 Specific Details

The three methods presented here go in order from how online they are. We start with the full batch approach, which takes all the data that needs to be segmented at once.

#### 3.1 Full Batch

##### 3.1.1 Finding Boundaries

Enhanced k-means cluster, just find the centers LULERINO KAPPUCINO.

#### 3.2 Mini Batch

Take fixed window sizes, find one boundary, work from there.

#### 3.3 Full Online

Essentially minibatch, but happy once a certain quality is reached.

#### 3.4 Parameters

Reducenumber, number of features used. Delta and amount of features used for distance matrix. Distance measure (dtw versus just standard versus lazy as fuck :D )

## **4 Results**

Compare with results from other papers.

### **4.1 Toydata**

### **4.2 Real data**

We use CMU motion capture data set.

## 5 Conclusion

Idk I don't think it works super well right now.

## References