# A literary review of machine learning for automatic classification of windmill turbines

Written by
Yohann Jacob Sandvik

**Abstract**

What happens if I make a change. Will it show up in the gitrepo?

# Contents

**1 Introduction** **2**

**2 Theory** **2**
2.1 Time series . . . . . 2
2.1.1 Auto-regressive moving average models (ARMA) models 2
2.1.2 Hidden Markov Models (HMMs) . . . . . 2
2.2 Time series clustering . . . . . 2
2.3 Neural networks . . . . . 3

**3 Condition based monitoring** **3**
3.1 Predictive maintenance . . . . . 3
3.2 Windmill turbines . . . . . 3
3.3 Treating sensor-data as a multivariate time series . . . . . 3

**4 Time series clustering - Shape-based approach** **3**
4.1 Model performance . . . . . 3
4.2 Cluster interpretation . . . . . 3
4.3 Time, and memory complexity . . . . . 4

**5 Time series clustering - Feature-based approach** **4**
5.1 Model performance . . . . . 4
5.2 Cluster interpretation . . . . . 4
5.3 Time, and memory complexity . . . . . 4

**6 Time series clustering - Model-based approach** **4**
6.1 Model performance . . . . . 5
6.2 Cluster interpretation . . . . . 5
6.3 Time, and memory complexity . . . . . 5

**7 Sub-sequence time series clustering** **5**

**8 Discussion** **5**

**9 Conclusion** **5**

# 1 Introduction

This section will be a general introduction to the assignment. Present the:

- Motivation
- Scope of the project assignment
- "Long term" goals of the project

Long term goals meaning the ability to predict whether a turbine needs maintenance before failure. To save the costs of regular maintenance and reduce down-time of windmills. As of now I was thinking that the scope of the project assignment could be a literary review of:

1. Clustering of time series data
2. Performance indicators of wind farms
3. Condition monitoring of wind turbines
4. Machine learning methods used for condition monitoring of wind turbines

# 2 Theory

## 2.1 Time series

Give a rigorous definition of a time series model. Describe the different domains of time series analysis, and time series forecasting. Introduce the data that we have as a multivariate time series model. Then continue introducing the time series models that I will be testing in the model-based clustering techniques.

### 2.1.1 Auto-regressive moving average models (ARMA) models

### 2.1.2 Hidden Markov Models (HMMs)

## 2.2 Time series clustering

Brief introduction to clustering techniques. Go through all the different approaches that can be made, such as representation methods, similarity measures and evaluation metrics.

### 2.3 Neural networks

Present neural networks. Explain how they usually are used as classifiers, but can be used as encoders to extract features from the time series.

## 3 Condition based monitoring

First give a short summary of the spectra of different condition monitoring schemes.

### 3.1 Predictive maintenance

Importance of predictive maintainance, and **short summary** of current implementations of predictive maintainance schemes.

### 3.2 Windmill turbines

Give a slightly more in-depth summary of current implementations of condition monitoring on windmill turbines

### 3.3 Treating sensor-data as a multivariate time series

Here I would introduce the magnitude of the amount of data produced by a single windmill turbine, and illustrate the necessity for tools that can handle this amount of data in real-time. Transition into the new section of clustering of time series.

## 4 Time series clustering - Shape-based approach

(Espens master)

### 4.1 Model performance

### 4.2 Cluster interpretation

Interpretation includes physical meaning of the different clusters, in regards to

- Fault diagnosis

- Performance measurement [1]
- What type of wind turbine it is (Siemens, Hydro?, etc.). Sounds simple, but even being able to detect this would make it much easier to detect which configurations they should "flash" each turbine with. Because now they have to check each turbine manually.
- Cluster affiliation of different turbines over time.

### 4.3 Time, and memory complexity

## 5 Time series clustering - Feature-based approach

Here I will expand upon the current implementations of time series clustering, that first perform some type of dimensionality reduction before clustering. Have found some examples of people using neural nets for feature extraction and then use clustering on medical time series.

### 5.1 Model performance

### 5.2 Cluster interpretation

### 5.3 Time, and memory complexity

## 6 Time series clustering - Model-based approach

In this approach they usually fit an ARMA / ARIMA model first, and then cluster the time series based on their model parameters

[1] Check the two papers talking about ML for measuring performance, and condition monitoring under performance based contract.

### 6.1 Model performance

### 6.2 Cluster interpretation

### 6.3 Time, and memory complexity

## 7 Sub-sequence time series clustering

Sub-sequence time series (STS) clustering has received a lot of negative publicity in the articles referenced in Espens master. But, I'm not sure how we are going to do without it, if the aim is to cluster time series in real time.

## 8 Discussion

## 9 Conclusion