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 Questions for Pierluigi and Gerthory **3**
1.1 Structural . . . 3

2 Introduction **3**
2.1 Motivation . . . 3
2.2 Objective . . . 3
2.3 Structure of this literary review . . . 4

3 Theory behind time series clustering **4**
3.1 Representation methods . . . 4
3.2 Time series modeling . . . 4
3.2.1 Auto-regressive moving average models . . . 4
3.2.2 Hidden markov models . . . 4
3.3 Similarity metrics . . . 4
3.4 Clustering algorithms . . . 4
3.5 Evaluation indices . . . 4
3.6 Feature extraction . . . 4

4 Basics of wind turbine condition monitoring **4**
4.1 Wind turbine components . . . 5
4.2 Treating sensor-data as a multivariate time series . . . 5

5 Time series clustering - Shape-based approach **5**
5.1 Model performance . . . 5
5.2 Cluster interpretation . . . 5
5.3 Time, and memory complexity . . . 6

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

7 Time series clustering - Model-based approach **6**
7.1 Model performance . . . 6
7.2 Cluster interpretation . . . 6
7.3 Time, and memory complexity . . . 6

8 Sub-sequence time series clustering **6**

9 Discussion **6**

10 Conclusion **6**

1 Questions for Pierluigi and Gerthory

1.1 Structural

- Should the first letter in every word of the sections be capitol? What about section 3.2.2

2 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

2.1 Motivation

2.2 Objective

A literary review of work using time series clustering for condition monitoring of wind turbines.

2.3 Structure of this literary review

3 Theory behind time series clustering

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.

3.1 Representation methods

Present the different ways that times series can be represented, time-domain, DFT, Wavelet transform, etc. Transition into time series models by ending section with the fact that time series can be modeled by making assumptions about how data was generated creating **models**.

3.2 Time series modeling

3.2.1 Auto-regressive moving average models

3.2.2 Hidden markov models

3.3 Similarity metrics

3.4 Clustering algorithms

Brief introduction to clustering techniques. Go through how the different algorithms cluster time series based on *similarity*

3.5 Evaluation indices

3.6 Feature extraction

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

4 Basics of wind turbine condition monitoring

First give a short summary of the spectra of different condition monitoring schemes. Importance of predictive maintainance, and **short summary** of

current implementations of predictive maintainance schemes.

4.1 Wind turbine components

Introduce the basic components of a wind turbine, and present some fun statistics about when the different components fail.

4.2 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.

5 Time series clustering - Shape-based approach

(Espens master)

5.1 Model performance

5.2 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.

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

5.3 Time, and memory complexity

6 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.

6.1 Model performance

6.2 Cluster interpretation

6.3 Time, and memory complexity

7 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

7.1 Model performance

7.2 Cluster interpretation

7.3 Time, and memory complexity

8 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.

9 Discussion

10 Conclusion