A Literature Review of Time-Series Clustering Techniques and Machine Learning Techniques Used for Monitoring of Wind Turbines

Written by
Yohann Jacob Sandvik

Abstract

Here the abstract will be.

Contents

Chapter 1

Introduction

Clustering is a method used to categorize big amounts of data into groups known as *clusters*, when there is little, or no information available about the underlying groups. It is a popular choice for extracting patterns from large datasets, because clustering falls within the category of *unsupervised machine learning*, meaning that it does not require the dataset to be labelled. Clustering is also a common step in data mining algorithms, where the goal is to learn rules relating the different variables in a dataset. In the last decade clustering has started to become more common to use on time-series datasets as they are abundunt, and labeling is often costly and time-consuming. Time-series clustering has been applied on financial time series, medical time series and time series from a variety of other industries.

1.1 Motivation

As of 2018 wind power, together with solar power made up 7% of the worlds electricity production,[1] and has been referred to as "the fastest growing source of energy" by the Norwegian company Statkraft[2]. As the effects of climate change steadily are becoming a reality shifting to renewable energy sources is imperative, and wind power will certainly play a bigger part in meeting the worlds energy demand in the future.

To make wind power as a whole more lucrative, a good start would be to reduce the downtime, and improve the performance of turbines. The argument that time-series clustering may be a good approach for this is two-fold.

1. A single wind turbine can have several houndred sensors sampling up to every second, meaning that a wind farm can produce colossal amounts of time-series data. An unsupervised approach is usefull simply because labelling of all this data is cumbersome.
2. When wind farms become big enough it will become to costly to manually inspect every turbine to construct an effective model for condition monitoring, further automation is required [1]. Time-series clustering is then a good alternative for condition monitoring.

1.2 Objective

This literature review has three objectives in the form of questions.

[1] https://www.iea.org/geco/electricity/

[2] https://www.statkraft.com/globalassets/old-contains-the-old-folder-structure/documents/wind-power-aug-2010-eng_tcm9-11473.pdf

Objectives

1. What machine learning methods are currently being used to monitor the condition, and performance of wind turbines?
2. What are the different methods of model based time-series clustering currently used?
3. What time-series clustering methods (if any) are appropriate to test on time-series data produced by wind turbines?

The literature review is meant to be a preliminary work for a master thesis where select techniques will be evaluated on actual time series data produced by a wind farm in Norway. The project assigment is also a continuation of the master thesis written in the spring of 2019 by Espen Waaga. In his thesis he explored the effectiveness of clustering raw time series in regards to similarity in time, and shape. In his "Future work" section Espen Waaga suggests that a natural next step would be to look at clustering with regards to similarity in change, which is done by implementing a model-based approach.

1.3 Structure of Review

Chapter 2

Method

2.1 Search terms

To find the relevant literature on the subjects of interest the search engine Oria was used to search the university library of the NTNU. Oria was preferred over other search engines such as Google Scholar because Oria allows one to combine multiple search terms in unison using "AND" or "OR", and because it allows the user to choose whether the search term should be in the title, subject, or other parts of the articles. The review will only consider articles published in peer-reviewed journals. Table 2.1 summarizes the search results. The *Title* and *General content* columns show which terms were used in the different searches; which terms where required to be in the title, and which terms could be in the "general content", meaning any part of the article. Let "$\times$" represent the AND operator between two search terms, and "$\wedge$" represent the OR operator. The "*" operator means that the search will include any word beginning with the word before the star, e.g *detect** includes *detection*, *detecting*, *detected*, etc. The N_f and N_i columns show how many results each search yielded and how many articles from each search were included in the review, respectively.

Nr.	Title terms	General terms	N_r	N_i
1	time $\times$ series $\times$ clustering	None	219	121
2	(monitor* $\wedge$ detect*) $\times$ clustering	time $\times$ series	187	21
3	wind $\times$ turbine* $\times$ (monitor* $\wedge$ detect*) $\times$ review	None	32	3
4	wind $\times$ turbine* $\times$ (monitor* $\wedge$ detect*)	machine $\times$ learning	100	47
		Total number of articles included		193

Table 2.1: Search results

2.2 Screening method

To make sure that the articles used were relevant, the review is limited to articles published 2014 or later. Some older articles are included through backward snowballing for their historical importance. There were three levels of screening, screening of the title, abstract, and full article. Title-screening was primarily for seeding out duplicate articles returned from the search-engine. The screening of the abstract and full-article were to identify the articles that were not relevant for the review and exclude them.

It has been a challenge to include enough literature to get a good overview of the different methods within time series clustering, but also not more literature the author alone could handle in the time available. So although the objectives is to get an overview of the different time series clustering methods, and an overview of the current machine learning methods used for monitoring wind turbines, the author does not claim to have made a complete exhaustive summary of all the possible methods.

When screening articles from search one and two, articles meeting one (or more) of the following criteria were discarded:

- Primary goal is time-series forecasting. It is outside the scope of this assignment.
- Time-series clustering is used only as a minor preprocessing step. Not considered relevant enough to the objectives of the review.
- Data used is not time series data. Not considered relevant enough to the objectives of the review.
- Paper does not actually use clustering algorithms. Not considered relevant enough to the objectives of the review.
- The data used consists of image time series. Data to different from data to be used in master thesis.
- The primary method is based on clustering raw time series wrt. similarity in time or shape, without transforming the representation method. This topic has been somewhat covered in Espen Waaga's master thesis, and is thus not considered relevant for this review.

Search number three was used to find existing literature reviews on condition monitoring of wind turbines. Three good literature reviews on the subject where found, and one good review on machine learning methods used for condition monitoring of wind turbines was found in search four. So, when screening the remaining articles from search number four the focus was to find articles not included in the aforementioned reviews, to complement them as well as possible.

Chapter 3

Time Series Clustering

Time-series clustering (TSC) can be divided into three components. How the time series is represented, which metric is used to measure similarity between time series, and what algorithm is used to cluster a set of time series. This chapter will give a short description of what specific type of TSC that this report will consider, it will go through the aforementioned steps of a TSC system, give an in-depth description of some time-series models and finally show some metrics used to evaluate a TSC system.

In this report we will deal with *discrete time series*. A time series is defined as a set of observations $\{x_t\}$ recorded at a specific time t. A discrete time series is a time series where the set of times when observations are made (T_0) is discrete [2]. A multivariate time series can be viewed as a set of vectors $\{\mathbf{x}_t\}$ where each set of vector elements $\{x_t^i\}$ is an individual time series. This means that the elements of the same vector $[x_t^1, x_t^2, ..., x_t^N]$ are separate observations made at the same time instance t. In a wind turbine, measurements of the temperature in the gear box made every 10 seconds can be considered a univariate time series. While the set of measurements made every 10 seconds of the temperature in the gear box, the power produced by the turbine, and the wind speed ahead of the blades can be considered a multivariate time series.

3.1 Types of Times-Series Clustering

There are three types of TSC, *whole-series TSC*, *subsequence TSC* and *time-point TSC*. Whole-series TSC is when multiple "whole" time series are clustered with respect to each other. Subsequence TSC comprises the clustering of subsequences of the same time series with respect to each other. The defining difference between whole-series and subsequence TSC is that whole-series TSC clusters multiple time series while subsequence TSC clusters clusters different subsequences of the same time series. In time-point clustering one clusters individual data points from multiple time-series with the same timestamp with respect to each other. One can consider it as regular clustering on the datapoints of a time series at a single time-instance. In this review we will mostly consider work using whole-series TSC, so unless specified one can assume that whole-series TSC is what is being refered to.

3.2 Representation Methods

There are numerous ways which a time series can be represented. Aghabozorgi, Shirkhorshidi, and Wah [3] define a time series representation given time series data $\{x_t\} = \{x_1, x_2, ..., x_T\}$ as transforming the time series into another vector $\{x_t\} = \{\hat{x}_1, \hat{x}_2, ..., \hat{x}_L\}$ where $L < T$. In theory $L = T$, but most often the point in transforming a time series is to reduce the amount of information present in the raw time series to more easily reveal patterns of interest. According to Aghabozorgi, Shirkhorshidi, and Wah [3] and Philippe Esling [4] representation methods can broadly be categorized into four categories:

- **Non-data adaptive**: Here the parameters of the transformation are the same for every time series [4]. Spectral transformations such as the Discrete Fourier Transform (DFT) and the Discrete Wavelet Transform (DWT) fall within this category.
- **Data adaptive**: The name here implies that the parameters of the transformation such as the length of the time series can change for every specific time series [4]. These methods focus on approximating a given time series in the best possible manner by reducing the global reconstruction error [3].
- **Model based**: These representations attempt to fit a stochastic model to the data [3]. Examples of models include Autoregressive Moving Average (ARMA) models, and Hidden Markov Models (HMM).
- **Data dictated**: This representation method differs from the aforementioned methods in as much that it automatically chooses the compression ratio based on the characteristics of the raw time series [3].

Data adaptive representations will be better at approximating time series than non-adaptive methods, but since transformation parameters can change for different time series it is harder to compare time series [3]. What is worth noting about model based representation methods is that they make assumptions about the underlying process that is generating the time series data [4]. Hence, it can be a good way of integrating a priori knowledge about the time series into the clustering system. In the sub-section below there is a more in-depth description of the two common time series models ARMA models, and HMM.

3.3 Time Series Models

To select suitable mathematical models for a dataset, we have to allow for the random nature of future observations. This is done by assuming that each observation in a time series x_t is a realization of a particular random variable X_t. The time series can then be modelled as a collection/set of random variables $\{X_t\}$, also known as a *stochastic process* [2].

To define an ARMA model, one needs to have a clear understanding of the terms white noise process, and stationary process. We say that the stochastic process $\{Z_t\}$ is "white noise" with zero mean, and variance σ^2 ($\{Z_t\} \sim WN(0, \sigma^2)$) if and only if $\{Z_t\}$ is zero mean, and every random variable contained in $\{Z_t\}$ is uncorrelated with every other random variable contained in $\{Z_t\}$. A stochastic process $\{X_t\}$ is said to be weakly wide-sense stationary if the mean, and variance are constant for all terms in the process. $\{X_t\}$ is said to be weakly short-term stationary if the mean and variance of terms are constant for distinct time periods within the duration of the process, but are not constant for all terms in the process. For brevity the term "stationary process" will be used when refering to a *weakly wide-sense stationary process.*

3.3.1 Autoregressive Moving Average Models

An ARMA model descirbes a time series in terms of difference equations. It can be considered a combination of two smaller models, an autoregressive (AR) model and a moving average (MA) model. Let $\{X_t\}$ be a stationary process. An MA(q) model will describe every term X_t as a linear combination of q distinct white noise terms as in equation (3.1).

$$X_t = Z_t + \theta_1 Z_{t-1} + ... + \theta_q Z_{t-q} \tag{3.1}$$

Whereas an AR(p) model will describe every term X_t as a linear combination of p previous terms of $\{X_t\}$ as in equation (3.2)

$$X_t = \phi_1 X_{t-1} + ... + \phi_p X_{t-p} \tag{3.2}$$

Putting equations (3.2) and (3.1) together, an ARMA(p, q) model will describe every term X_t as a linear combination of p previous terms, and q white noise terms as in equation (3.3).

$$X_t - \phi_1 X_{t-1} - ... - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + ... + \theta_q Z_{t-q} \tag{3.3}$$

Given that the polynomials $1 + \theta_1 z + \theta_2 z^2 + ... + \theta_q z^q$ and $1 - \phi_1 z - \phi_2 z^2 - ... - \phi_p z^p$ have no common factors [5].

3.3.2 Hidden Markov Models

Let $\{X_n\}$ be a stochastic process where the random variables contained in $\{X_n\}$ only can take on a finite number of values which we will call states. Let X_n denote the state at time period n. The probability of X_n transitioning from state i to state j at the next time period $n + 1$ is denoted p_{ij}. It might seem natural that p_{ij} is conditional on what the state was in the last time period. $\{X_t\}$ is said to be a *Markov chain* if p_{ij} only is conditional on the last past state, as shown in equation (3.4)

$$\begin{aligned} p_{ij} &= P(X_n = i | X_{n-1} = i_{n-1}, X_{n-2} = i_{n-2}, ..., X_1 = i_1, X_0 = i_0) \\ &= P(X_n = i | X_{n-1} = i_{n-1}) \end{aligned} \tag{3.4}$$

Suppose now that the states that the process is in are hidden from the observer. Instead there exists a finite set of signals $\{S\}$ that are emmitted when the process enters a state. In addition, let the probability of emmitting signal s, at time period n, in state j ($P(S_n = s|X_n = j)$) be independent of previous states, and signals emmitted. A model of this type where the signals $S_1, S_2, ...$ are observed, and the underlying Markov states remain hidden is called a *hidden Markov chain model* [6].

3.4 Similarity Metrics

One similarity metric is euclidean distance.

3.5 Clustering Algorithms

As mentioned before clustering is a form of unsupervised machine learning. The goal is to divide the dataset into clusters, by maximizing some similarity metric for members of the same cluster, and minimizing the same metric for members of different clusters.

3.6 Evaluation Indices

Chapter 4

Appendices

4.1 Time-series clustering

Ref.	Representation	Similarity measure	Clustering Algorithms	Evaluation
[7]	Vibration signal.	ARMA and J48	Tests.	Evaluation
[8]				

Table 4.1: Summary of model based time-series clustering methods

Bibliography

[1] Espen Waaga. *Machine Learning for Automatic Classification of Wind Turbines*. English. Master thesis. 2019.

[2] Peter J. Brockwell and Richard A. Davis. *Time Series: Theory and Methods Second Edition*. Springer-Verlag New York, Inc., 1991.

[3] Saeed Aghabozorgi, Ali Seyed Shirkhorshidi, and Teh Ying Wah. "Time-series clustering - A decade review". eng. In: *Information Systems* 53 (2015), p. 16. ISSN: 0306-4379.

[4] Carlos Agon Philippe Esling. "Time-series data mining". In: *ACM Cumputing Surveys* 45 (2012).

[5] Peter J. Brockwell and Richard A. Davis. *Introduction to Time Series and Forecasting Third Edition*. Springer International Publishing Switzerland, 2016.

[6] Sheldon M. Ross. *Introduction to Probability Models 11th Edition*. Academic Press Elsevier Inc., 2014.

[7] A. Joshuva; V. Sugumaran. "A machine learning approach for condition monitoring of wind turbine blade using autoregressive moving average (ARMA) features through vibration signals: a comparative study". English. In: *Progress in Industrial Ecology, An Int. J.* 12.1/2 (2018). ISSN: 1476-8917. URL: `http://www.inderscience.com/link.php?id=95867`.

[8] P. Qian et al. "A novel condition monitoring method of wind turbines based on long short-term memory neural network". English. In: *Energies* 12.18 (2019). ISSN: 19961073.