

A LITERATURE REVIEW OF TIME-SERIES CLUSTERING TECHNIQUES AND MACHINE LEARNING TECHNIQUES USED FOR MONITORING OF WIND TURBINES

Yohann Jacob Sandvik
Institute of Electronic Systems - NTNU

Motivation

To make wind power as a whole more lucrative, a good start would be to reduce the down-time, 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 sensors sampling very often, meaning that a wind farm can produce colossal amounts of time-series data. An unsupervised approach is useful because labelling of all this data is requires a lot of time and resources.
2. When wind farms become big enough it will become too costly to manually inspect every turbine to construct an effective model for condition monitoring, further automation is required. time-series clustering is therefore a good alternative for condition monitoring.

Objectives

- 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 feature-, and model-based TSC currently used?
 3. What TSC methods are appropriate to test on time-series data produced by wind turbines?

Search terms used to find literature

Nr.	Title terms	General terms	N_r	N_i
1	wind × turbine* × (monitor* ∧ detect*) × review	None	32	3
2	wind × turbine* × (monitor* ∧ detect*)	machine × learning	100	47
3	time × series × clustering	None	219	46
Total number of articles included			96	

Search term explanation

- The search engine Oria was used to search the university library of the NTNU.
- 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.
- "×" represents the AND operator, and "∧" represents the OR operator between search terms.
- The "*" operator means that the search will include any word beginning with the word before the star. For example detect* includes words such as detection, detecting and detected.
- 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.
- Search 1 and 2 are used to find articles covering the first objective and search 3 is ment to cover the second objective.

Screening method

- To make sure that the articles used were relevant, the review is limited to articles published in peer-reviewd journals, after January 2014.
- Search number 1 was used to find existing literature reviews on condition monitoring of wind turbines. Three good literature reviews on the subject were found so when screening the remaining articles from search 3 the focus was to find articles not included in the aforementioned reviews.
- When screening articles from search 3 articles meeting one (or more) of the criteria outlined in the table below were excluded from the review.

Criteria	Reason
Primary goal is time-series prediction or time-series forecasting.	It is outside the scope of this assignment.
Uses the subsequence time-series clustering method.	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 time-series clustering methods explored in the work are not model-based or feature-based.	The raw-data-based approach has been somewhat covered in Espen Waaga's master thesis, hence it is emmitted in this review.
The specific model-based approach using the tail dependence of time series.	Method not found relevant enough for the data that will be used in the master thesis, more relevant for financial time series.