Machine Learning Methods for Wind Turbine Condition Monitoring

Srikar Lankamalla - 2018A4PS1035H
Vishwanath Guruvayur - 2018A4PS1048H
Niladri Nilamadhab - 2018A4PS0229H
Akhil Annapaneni - 2018A4PS1213H
Ackashh Dhevaraju - 2018A4PS1210H
Ankit Kaviath - 2018A4PS1208H

Overview

- Condition Monitoring
- Machine Learning
- Data Set, Feature Selection
- Regression Based Analysis
- Classification based Analysis
- Validation, ML models in CM decision support systems

Condition Monitoring

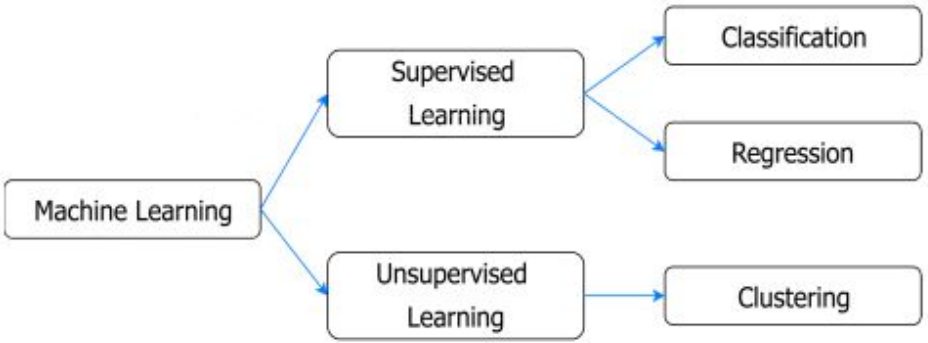
- Condition Monitoring is a part of Operations and Management. This includes management, control and maintenance which is done in order to keep the wind farm running.
- We have different levels at which we can perform Condition Monitoring. At a lower level, we can monitor the condition of different sub-components of the turbine and at a higher level, we can monitor the condition of the wind farm as a whole.
- We also have two main types of monitoring. **Intrusive Monitoring** - comprises of methods like vibration analysis, shock pulse methods, oil debris monitoring ,etc. **Non-intrusive monitoring** - comprises of methods like visual inspection,ultrasonic techniques, etc.
- We can use Condition Monitoring to identify faults in the future or even in real-time scenarios. **CM for diagnosis** - fault detection in present timeline. **CM for prognosis** - fault prediction for future timeline.
- Different components have different failure rates and properties. This must be considered during Condition Monitoring and priority must be given accordingly.

Machine Learning Overview

The process of developing a model that gains information from data without the assistance of an expert is referred to as machine learning (ML).

A series of steps can be used to represent the ML process:

- Data collection and preprocessing
- Feature selection and extraction
- Choosing a ML model
- Validation of the Model



- Neural Networks (NN) and Support vector machines (SVM) are two popular ML methods for diagnostics and prognostics.
- In the wind energy sector, NNs are being extensively used for predicting (e.g.,forecasting of wind velocity), control (e.g., wind turbine power control), recognition,& assessment(e.g. Diagnosis of faults.)
- Support Vector Machines are frequently used in detection of faults and, more broadly, for complex data sets.
- The validity of the generated Machine Learning models can be estimated using a range of specific methods in conjunction with techniques like k-fold cross validation method, which determines how accurately the model's output will make assumptions with data different from the training data.

Data Set, Feature Selection

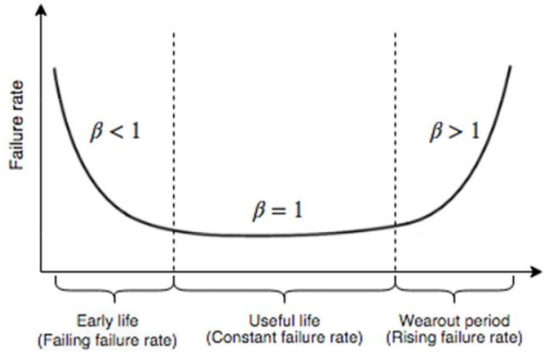
- Wind farms generally require a couple of specific type of data points such as:-
  - **Volume:-** Amount of data produced. Generally around 200 MB per second per turbine
  - **Velocity:-** Frequency at which data is produced
  - **Variety:-** Images, Video data and action data
  - **Veracity:-** Data provided in above cases shouldn't be inconsistent, or more data would be required
- Feature Selection:-
  - This process involves selecting the specific variables and data which we want to study

- Methods of feature selection:-
  - Wrapper:- Views Machine Learning algorithms as small black boxes and provides different subsets of features required for output.
  - Filter:- Conducts tests between feature and signal and rank them accordingly.
  - Embedded:- Used in combination with the filter method. Considers the relative importance of features per node.
- Feature Extraction and Methods:-
  - Feature extraction is utilized to compress high-dimensional time series while making sure that the characteristics do not change. This has the effect of speeding up the model training.
    - Statistics:- Simplest and Most efficient
    - Parameters of Fitted Models:- Coefficient of Fitted models, and Auto-Correlation statistics.
    - Time-frequency Properties:- Transform the time domain based signals into the frequency domain.

Regression Based Analysis

Regression Based Analysis is mainly carried out in order to detect any anomalies in the Wind Turbine's Performance.

This is done by first, modelling the regression model for the turbine when it is in it's peak performance i.e. in the healthy state by using the various individual input parameters like wind speed, pressure, etc. and output parameters like Power, Force, Torque, etc.

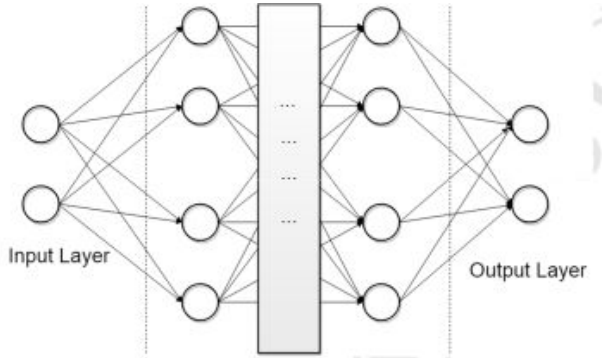


After developing the regression model, when the new input data from the current Wind Turbine under analysis comes in, this data is used to find the expected value of the output parameters according to the developed regression model.
If the current data is seen to differ by a significant amount, an alarm is raised so that the components related to that specific input parameter can be checked for any possible errors.
These regression models at the normal/optimal performance can be generated with many different levels of complexity.
Eg: For modelling the Power Curve which is the relation between Power generated by the wind turbine v/s the Wind Speed at the location; is considered to be conceptually of the highest level as the entire Wind Turbine model is considered to be a black box.

Regression Based Analysis

There are 2 techniques of modelling Power curves:
A] **Parametric Modelling Techniques** : Models are constructed by considering Parametric Vector made up of a finite set of Parameters.
For Eg: Linear Segmented Model (2), 4-Parameter Logistic Function, Polynomial Curves, etc.

B] **Non-Parametric Modelling Techniques**: Models are constructed by Parametric Vectors that are not bounded wrt length.For Eg: Neural Networks, Fuzzy Logic Methods, Cubic Spline Interpolation, etc.



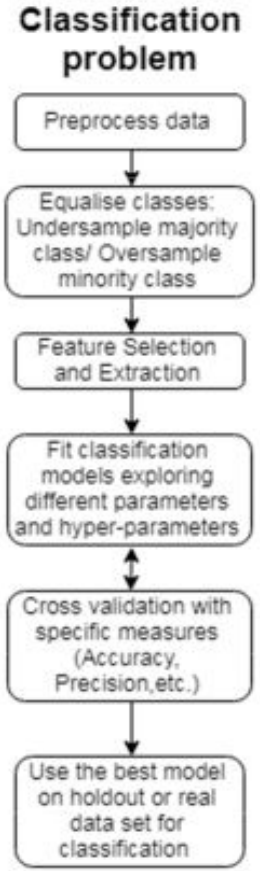
Power curves are generally provided by wind turbine manufacturers. These models are specifically made with respect to the location where the company's turbines were put to test, which means they were in accordance to the location specific meteorological conditions which will differ to the conditions in the location in which the Turbine will be installed.
We will be able to efficiently assess and predict the potential of wind turbine power at the site and plan for more efficient wind turbine modelling choices if we empirically model these power curves independently.

Classification based Analysis

Classification based analysis involves the finding of a relationship between various different input variables which are generally represented in the form of vectors and one of categorical variable (output) is identified by labels.

While training the model, the model is iteratively given an input vector of variables along with the label/class of the state. The input vector of variables might comprise of various features taken from pre-processed timed series of signals which will be in relevance to the component modelled.
For example: In the case of Condition Monitoring of a Generator of a Wind Turbine we can have output failure labels such as healthy/brush failure/winding failure/ etc.

Once the model is trained with appropriate amount of data, the model is used as a prediction model for the Condition Monitoring of Wind Turbine Models based on the various parameters at the location of establishment as a function of time.



Classification based Analysis

As this method comes under supervised Machine Learning, the output labels/classes to be given signifying to exactly which training data points they belong to. This is a highly time consuming process and it is prone to a lot of errors because of which there is a probability of resulting in input vectors with an imbalance in the number of classes/labels. This is a very frequently occuring issue when performing classification type analysis.

Ways of getting rid of such problems is to undergo methods like:

- Under-Sampling : We have to exclude data-points that belong to the majority class/label.
- Oversampling: We have to include more data-points from minority class/label.
- Tomek-links: We have to remove such data-points in the majority class which are termed as redundant/ noise/ borderline/ etc.
-

Most commonly used methods are Support Vector Machines (SVM), Boosting Tree Algorithm (BTA), etc.

Validation

- One of the most popular methods followed for validation is “K- fold cross validation”.
- This involves splitting the whole data set into k-bins(typically k=10), then selecting one random bin for test purpose and the other k-1 bins are used for train purpose.
- This process is run k times and the scores are averaged to tell the performance of a specific model.

For Validity of regression-based normal behaviour models, performance measures are:

- MAE and MAPE for measuring error in prediction.
- sMAPE is for measuring absolute error in prediction.
- RMSE for explaining spread in the prediction.
- R^2 to explain how well the prediction is made.

For classification-based models:

- the number of True Positives, True Negatives, False Positives and False Negatives are considered and
- Measures such as Accuracy,Recall,Specificity, precision are evaluated.

Integrating ML CM models for decision support

- Finally, integrating these developed ML models for real time information will help in decision making for the operators.
- ML CM models also provide an advantage of looking over how are the conclusions reached.
- There are some of the “White Box” models such as Decision Trees which can be easily interpreted and there are also some “Black Box” models such as Neural Networks which provide no information.
- Appending confidence levels for the predicted values can help in decision strategy. Though some models like Logistic Regression are probabilistic , it can also be added for some other models.
- Sensors collecting data also must be regularly updated as this may lead to degradation in performance measures.

Conclusion

Learnings from the term paper:

1. Machine learning models for Condition Monitoring
2. Regression based, classification based models
3. Validation and their integration