The Impact of Feature Causality on Normal Behaviour Models for SCADA-based Wind Turbine Fault Detection

Anonymous Authors[1]

Abstract

The cost of wind energy can be reduced by using SCADA data to detect faults in wind turbine components. Normal behavior models are one of the main fault detection approaches, but there is a lack of work in how different input features affect the results. In this work, a new taxonomy based on the causal relations between the input features and the target is presented. Based on this taxonomy, the impact of different input feature configurations on the modelling and fault detection performance is evaluated. To this end, a framework that formulates the detection of faults as a classification problem is also presented.

1. Introduction

In 2018, global energy-related CO_2 emissions reached a historic high of 33.1 gigatonnes. These emissions are caused by the burning of fossil fuels, mainly natural gas, coal and oil, which accounted for 64% of global electricity production in this same year (IEA, 2018). Greenhouse gases like CO_2 are responsible for climate change which threatens to change the way we have come to know Earth and human life. For the previous reasons, there has been a global effort to shift from a fossil fuel based energy system towards a renewable energy one. In fact, it is expected that by 2050 wind energy will represent 14% of the world's total primary energy supply (DNV-GL, 2018).

The operation and maintenance costs of Wind Turbines (WTs) can account for up to 30% of the cost of wind energy (EWEA, 2009). This happens because while generators in fossil fuel power plants operate in a constant, narrow range of speeds, WTs are designed to operate under a wide range of wind speeds and weather conditions. This means that stresses on components are significantly higher, which increases the number of failures and consequently the maintenance costs. There have been recent efforts to monitor and detect incipient faults in WTs by harvesting the high amounts of data already generated by their Supervisory Control and Data Acquisition (SCADA) systems, which, in turn, enables the wind farm owners to employ a predictive maintenance strategy. In fact, it is expected that by 2025 new predictive maintenance strategies can reduce the cost of wind energy by as much as 25% (IRENA, 2016). One of the main methods for monitoring the condition of WTs is building Normal Behaviour Models (NBMs) of the component temperatures. The fundamental assumption behind the use of NBMs is that a fault condition is normally characterized by a loss of efficiency, which results in increased temperatures. By using SCADA data to build a model of the temperatures of the WT components, one can calculate the residuals, which are the difference between the real values measured by the sensors and the predicted values by the model. These residuals can be used to detect abnormally high temperatures that may be indicative of an incipient fault. Multiple works (Zaher et al., 2009; Mesquita et al., 2012; Brandao et al., 2015) have reported good results using NBMs to predict WT failures, being able to predict failures in WT components months in advance. In these works the authors used as features Active Power, Nacelle Temperature and lagged values of the target temperature, thus including autoregressive properties into to the model, to predict the temperatures of various components. In (Schlechtingen & Santos, 2011) and (Bach-Andersen et al., 2016) the authors obtained an important result: although the use of autoregressive features resulted in better modelling performance it also resulted in worse fault detection performance. Another important result was obtained in (Bangalore et al., 2017) and (Tautz-Weinert, 2018), which indicated that using features that are highly correlated with the target also increased the modelling performance but decreased the fault detection performance of the model. The actual reason for such behavior is the use of features that are casually dependent on the target, such as temperature sensors physically close to the one being modelled.

In the present work we hypothesize that if there is an increase in the temperature of a faulty component, the physically close components will also get hotter due to heat

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

transfer. This means that using these physically close components as input features to the model may leak information regarding the fault state of the component, making the model unable to detect abnormal behaviour. Even though one can intuitively understand why these type of features should not be used for NBMs, it has not been shown in a consistent case study. In fact, most of the literature today still uses this type of features, examples include: (Colone et al., 2018; Tautz-Weinert & Watson, 2016; Cui et al., 2018; Mazidi et al., 2017; Zhao et al., 2018; 2017). This may be due to the lack of a clearer nomenclature on the types of input features and also due to the lack of comparisons between the different possibilities. For this reason, this work presents a new taxonomy based on Econometric Causality (Heckman, 2008), which distinguishes features based on their causal relations with the target. If the target is causally dependent of the features, these are Causal Features. On the other hand, if the target depends on the features but the features also depend on the target they are Simultaneity Features. Such causal relations are assumed based on the domain knowledge of the physical system. In this work, the impact of different input feature configurations, including autoregressive features, on the normal behaviour modelling and fault detection performances of the model will be compared.

It is also important to note that evaluating the fault detection performance of different models is not as trivial as evaluating their normal behaviour modelling performance. In fact, there is no standard in the literature regarding how to evaluate fault detection performance. This happens because of the inherent nature of the fault detection problem, in which there is rarely groundtruth. Indeed, there is data of when the failure happened, but there is no information regarding when the fault state started, making it not trivial to formulate as a classification problem. Hence why the majority of the literature evaluates the fault detection results by visual inspection, observing the increase in the residuals before the failure. This is problematic, because comparisons between different models will be highly subjective. Having this in mind, this work will also present a formulation of the detection of faults as a classification problem.

2. Methods

2.1. Data, Training and Residuals

The data used in this work comes from a wind farm composed of 15 turbines during a period of 6 years. The collected data correspond to SCADA signals with 10 minute resolution. During the year of 2012 there was a total of 5 failures in the drivetrain of the WTs, related with the Gearbox IMS Bearing. For these reasons, a NBM of the Gearbox IMS Bearing temperature will be trained for each WT, with the objective of predicting the corresponding failures.

The models will be trained with data from the beginning of 2007 to the end of 2011 and tested on data from 2012. It is also important to note that known periods with faults will be removed from the training data so the model does not learn abnormal behaviour. The models will be implemented with Gradient Boosting Decision Trees (GBDT), which work by iteratively combining weak decision trees into a strong ensemble learner. There are various open source implementations of GBDTs, in this work LightGBM (Ke et al., 2017) will be used due to is better computational performance. In terms of optimization, the year of 2011 will be used as a validation set when choosing the number of trees for each model by early stopping. Note that no exhaustive hyperparameter optimization was performed, so all models will use the same hyperparameters besides the number of trees.

2.2. Feature Configurations

Based on the taxonomy previously presented different models will be defined based on their input feature configuration. The simplest model that will be tested is the Causal Normal Behaviour Model (CNBM), which only uses causal features. These causal features are determined based on domain knowledge and will be: Rotor Speed, Active Power, Pitch Angle, Wind Speed and Ambient Temperature. All these features characterize the operation regimes of the WT, these are causal features because the Gearbox IMS Bearing Temperature depends on their values, but their values are not dependent on it. For example, variations in the Ambient Temperature influence the overall temperature of the WT and thus the Gearbox IMS Bearing Temperature, but the influences of the latter on the Ambient Temperature can be disregarded.

On the other hand, simultaneity features will be chosen based on Pearson Correlation, which is a standard first approach for regression problems. The highest correlated feature with the Gearbox IMS Bearing Temperature is the Gearbox HSS Bearing Temperature, which is a simultaneity feature because there is heat transfer between the two sensors, thus meaning that their values are mutually causally dependent. Having this in mind, the Simultaneous Normal Behaviour Model (SNBM) will use all the features from the CNBM plus Gearbox HSS Bearing Temperature. Two more models will be tested, which correspond to the autoregressive versions of the previously described models: Autoregressive Causal Normal Behaviour Model (ACNBM) and Autoregressive Simultaneous Normal Behaviour Model (ASNBM).

2.3. Fault Evaluation Framework

To develop an evaluation framework for fault detection, one must first formulate it as a binary classification problem

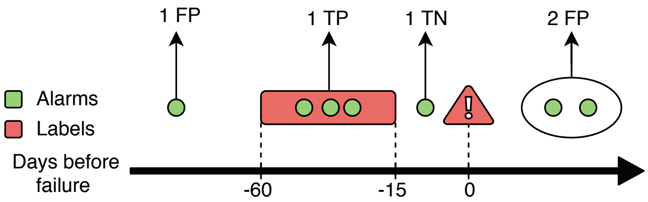

Figure 1. Schematic example of fault detection formulated as a classification problem.

where there are two labels: fault and no-fault. Since there is no information regarding the fault state of the component, only the date of failure, it was defined with the wind farm owners that for the failures studied in this work it can be assumed that a fault state would be present at most 60 days before the failure. Besides that, it was also defined that for the alarms to be useful they should be triggered at least 15 days before the failure. This means that to be considered a True Positive (TP) the alarm must be triggered between 60 and 15 days before the failure. In Figure 1 is presented a schematic example of the previously described fault detection classification problem formulation. Taking this example, it is important to note that the number of alarms triggered in the prediction window is not relevant, they are all aggregated as 1 TP. The main reason for this, is that if the aggregation is not done, then 4 alarms for the same failure would count as much as 4 detected failures with 1 alarm each. This clearly is not what is intended of the framework, since 1 alarm should be enough to motivate an inspection, and detecting 4 failures with 1 alarm weights more than detecting 1 failure with 4 alarms. Finally, it is also important to note that alarms triggered less than 15 days before the failure are not considered False Positives (FPs), since there is indeed a fault state, it simply is not relevant, so they are considered True Negatives (TNs).

3. Results

In terms of normal behaviour modelling, the models were evaluated on periods of turbines that are known to be healthy. The results, presented in Table 1, indicate that the use of simultaneity features indeed improves the modelling performance, since SNBM obtains better results than CNBM. The use of autoregressive features also improves the modelling performance, since ACNBM and ASNBM obtain better results than their non-autoregressive counterparts. This results make sense, since there are certain regimes of the turbine that are difficult to model without simultaneity nor autoregressive features, such as the turning off of the turbine as noted in (Bach-Andersen, 2017).

In terms of fault detection, a baseline was defined that consists of simply setting different thresholds on the distribution of the target temperature and obtaining the corresponding

Table 1. Regression error metrics for the training and test sets of each model.

Model	Training MAE	Training RMSE	Test MAE	Test RMSE
CNBM	1.48	2.14	1.80	2.62
SNBM	0.87	1.26	1.01	1.41
ACNBM	1.03	1.57	1.14	1.67
ASNBM	0.83	1.22	0.96	1.38

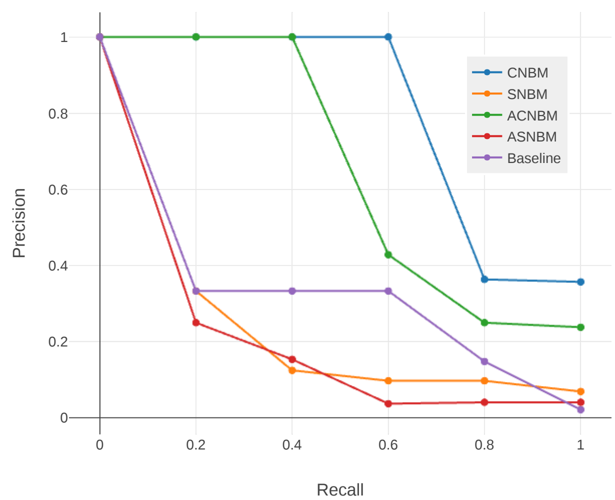

Figure 2. Precision and Recall curves for the different models.

precision and recall. For the models, also different thresholds were applied in the residuals to obtain the different values of precision and recall. The results are presented in Figure 2. As can be seen, the CNBM which obtained the worst modelling performance obtains the best fault detection performance. It is also important to note that the models with the simultaneity feature are significantly worse than the baseline.

4. Conclusions

An evaluation framework to formulate the detection of faults as a classification problem was presented, this hopes to contribute to the development of a standard approach for fault detection performance evaluation in the field. Besides that, a new taxonomy regarding the causal relations of the different input feature types was presented, which hopes to make the discussion on how different features affect the performance of models clearer. Finally, it was also demonstrated that although autoregressive and simultaneity features increase the modelling performance they decrease the fault detection capabilities of the model. This is an important contribution since the majority of works today still use these types of features.

References

Bach-Andersen, M. *A Diagnostic and Predictive Framework for Wind Turbine Drive Train Monitoring*. PhD thesis, Technical University of Denmark, 2017.

Bach-Andersen, M., Rmer-Odgaard, B., and Winther, O. Flexible non-linear predictive models for large-scale wind turbine diagnostics. *Wind Energy*, 2016.

Bangalore, P., Letzgus, S., Karlsson, D., and Patriksson, M. An artificial neural network based condition monitoring method for wind turbines, with application to the monitoring of the gearbox. *Wind Energy*, 2017.

Brandao, R., Carvalho, J., and Maciel-Barbosa, F. Intelligent system for fault detection in wind turbines gearbox. 2015.

Colone, L., Reder, M., Dimitrov, N., and Straub, D. Assessing the utility of early warning systems for detecting failures in major wind turbine components. *Journal of Physics: Conference Series*, 2018.

Cui, Y., Bangalore, P., and Bertling Tjernberg, L. An anomaly detection approach based on machine learning and scada data for condition monitoring of wind turbines. 2018.

DNV-GL. *Energy Transition Outlook*. DNV GL, 2018.

EWEA. *The Economics of Wind Energy*, 2009.

Heckman, J. J. Econometric causality. *International Statistical Review*, 2008.

IEA. *Global Energy & CO2 Status Report*. International Energy Agency, 2018.

IRENA. *The Power to Change: Solar and Wind Cost Reduction Potential to 2025*, 2016.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., and Liu, T.-Y. Lightgbm: A highly efficient gradient boosting decision tree. In *NIPS*, 2017.

Mazidi, P., Mian, D., Bertling, L., and Sanz Bobi, M. A. Health condition model for windturbine monitoring through neural networks and proportional hazard models. *Journal of Risk and Reliability*, 2017.

Mesquita, R., Carvalho, J., and Pires, F. Neural networks for condition monitoring of wind turbines gearbox. *J. Energy Power Eng.*, 2012.

Schlechtingen, M. and Santos, I. Comparative analysis of neural network and regression based condition monitoring approaches for wind turbine fault detection. *Mechanical Systems and Signal Processing*, 2011.

Tautz-Weinert, J. *Improved wind turbine monitoring using operational data*. PhD thesis, Loughborough University, 2018.

Tautz-Weinert, J. and Watson, S. J. Comparison of different modelling approaches of drive train temperature for the purposes of wind turbine failure detection. 2016.

Zaher, A., McArthur, S., Infield, D., and Patel, Y. Online wind turbine fault detection through automated scada data analysis. *Wind Energy*, 2009.

Zhao, H., Liu, H., Hu, W., and Yan, X. Anomaly detection and fault analysis of wind turbine components based on deep learning network. *Renewable Energy*, 2018.

Zhao, Y., Li, D., Dong, A., Kang, D., Lv, Q., and Shang, L. Fault prediction and diagnosis of wind turbine generators using scada data. *Energies*, 2017.