Study of Feature-Selection-Algorithms with powerful 3D insights for Wind Turbine Failure Prediction using SCADA Data

Jordi Solé-Casals[1*], Pere Marti-Puig[1], Juan José Cárdenas[2], Alejandro Blanco[2] M., Isaac Justicia[2], Jordi Cusidó[2]

[1] *Universitat de Vic, Vic, Spain*
[2] *ITESTIT S.L., Terrassa, Spain*
[*] *Presenting author*

Introduction

One of the possible techniques to use SCADA data in order to implement failure prediction is to use machine learning classifiers that can extract the patterns that precede a failure or the related fault-alarm. Among the different steps to carry out successfully the classifier implementation, one of the key steps is to choose optimal or near optimal set of inputs from tens or hundreds of variables that can feed the classifier. Fortunately, in machine learning, there are a lot of unsupervised-supervised techniques-algorithms that are aimed to find out the most relevant and shortest set of variables related with the respond variable. Therefore, with the aim of study and select the best algorithm or algorithms of feature selection when applied to wind turbine failure prognostic, in this paper we present a thorough study of the state of the art techniques for feature selection applied to the specific area of wind turbine Operation & Maintenance.

Although our studies confirm that a selected set of six to ten more discriminant variables are required to obtain the best prognosis performance, that selection is rather difficult to be represented. However, sets of three variables admit a 3D Cartesian plot and time evolution can be included by plot animation. Those dynamic representations provide powerful and intuitive insights about the behaviour of variables 21 days before failure and help us to improve the models used for prognostic. The examples presented in this work have been extracted from measured features taken along an entire year and correspond to gearbox and transmission systems of wind turbines, specifically for **Fuhländer** wind turbines. In addition, the best of input selection algorithms have been integrated to Smartive Cloud-based platform.

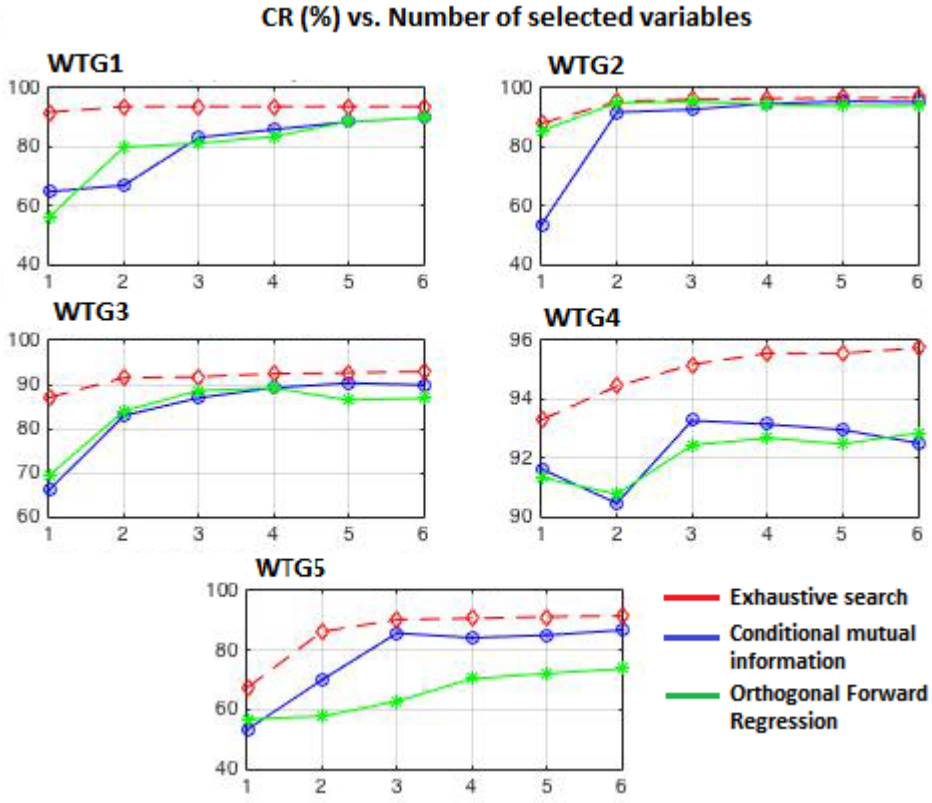
Approach

After a selection of state of the art algorithms for feature selection in classifiers, we have carried out a study of performance of those algorithms, building a KNN classifier with the selected features for each algorithm. The result of each algorithm is compared against the optimal result obtained by means of an exhaustive search algorithm. Using human expert knowledge, the set of 404 available variables from SCADA data has been reduced to a set of thirty in order to make viable the exhaustive search. The comparison is done by means of the classification rate (CR) of a KNN algorithm trained with the selected variables by

the feature selection algorithm under test. The CR is calculated as the average of 100 experiments using randomly selected observations from the main data set.

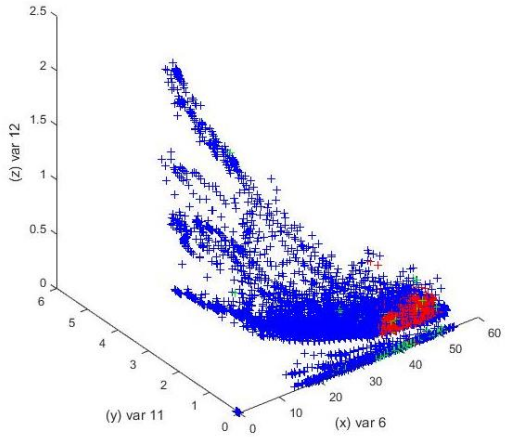
Main body of abstract

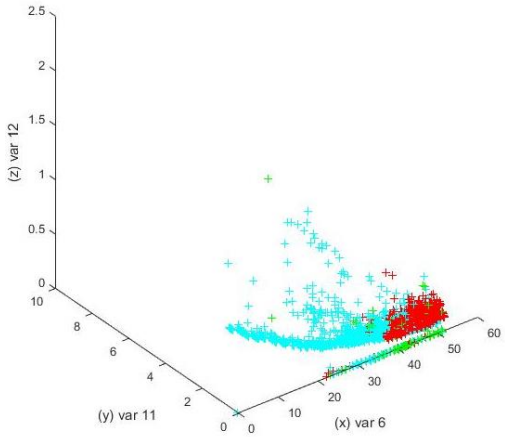
This methodology has been applied to five Fuhländer wind turbines, for data gathered in 2014. A total of 8 algorithms were tested, among them: conditional mutual information, orthogonal forward regression, minimum redundancy maximum relevance, conditional mutual information maximisation and joint mutual information. The next graphic shows two of the best obtained results for the five wind turbines and for 1 to 6 features to be selected.



In the graphs above we can see that in overall both compared methods have a good performance and are very close to the result obtained with the set selected with exhaustive search algorithm. In this sense, we can say that the automated methods (which are much faster) can help us to choose the best features, even from an bigger initial set of variables (the full 404 variables for example).

Finally, in the below graphic we can see a representation in 3D of a set of three selected features. In the left plot all data is plotted and in blue is the data with no alarm and in red with alarm. It is clear that a region of “alarm” or “failure” exist. In the right plot, just the data of some days before the alarm (light blue colour) is plotted together with the data when the alarm occurred (red colour). There, we can see that not just a region exist, but the values of the variables move to this region before the failure.





Conclusions

It has selected and assessed a set of algorithms for feature selection in machine learning classifiers when used for wind-turbine failure prediction. The best of them have been integrated to Smartive platform in order to support failure and health estimation modules. The obtained results show the viability of automating one of the most critic steps when building a classifier for failure prediction in wind turbines and reducing significantly the time of selection.

Learning objectives

- To know some of state of the art feature selection algorithms for machine learning classifiers.
- To learn how well those algorithms perform when used for classification and prediction of failures in wind turbines.
- To know that for some failures in wind turbines it is possible to find “failure” or “alarm” regions based on related variables from the SCADA data and that it is possible to anticipate the failure following the trend of the data when is near to those regions.