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[1,2] M., Isaac Justicia[2], Jordi Cusidó[2]

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

Abstract

One of the key steps in failure prediction using machine learning classifiers is to choose an optimal or near optimal set of inputs from tens to hundreds of variables. This task can be achieved with the implementation of unsupervised-supervised algorithms that aim to find out the most relevant and shortest set of variables related with the failure. Therefore, with the aim of study and select the best algorithm or algorithms of feature selection, we present a **thorough study of the state of the art of available techniques when applied to the specific area of wind turbine Operation & Maintenance**. In order to visualize the behavior of the selected variables we have choose sets of three variables for fault producing a understandable **3D plot** animations. Those provide **intuitive and powerful insights** about the behavior of the WTG until 21 days before failure. This helps us to confirm and improve the models used for failure prediction.

Introduction

Successful failure prediction from SCADA data requires a set of key steps and processing techniques to separate the noise from the important indicators. The noise is present in some level as non-relevant variables that must be removed in the first stages. This process, feature selection, reduces the number of input variables based on *target* state, from tens to hundreds, reducing the input variables dimension that contributes in reducing the computation requirements in time and space.

In the feature selection we can choose between supervised and unsupervised algorithms, this document covers a set of unsupervised feature selection algorithms applied to the specific area of wind turbine Operation & Maintenance.

The results of our studies confirm that a selected subset of at least 6 to 10 variable is enough to obtain the best prognosis performance that would be obtained from analyze all the possible variables with an exhaustive method. To visualize the results of the feature selection algorithms a set of 3D visualization of 3 variables from each selection results presents the variable behavior change since 21 days before the *target* failure, demonstrating the effectiveness of the feature selection.

These studies were applied specifically to the gearbox and transmission systems of a set of **Fuhländer's** brand wind turbines.

Objectives

During the development of the methodology and the evaluation of the results two main objectives where followed:

- Test how well those selection algorithms perform when used for classification and prediction of failures in wind turbines.

- Test if 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.

Methodology

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 in order to determine its classification performance in terms of the impact at the final results.

The process starts using human expert knowledge to select a subset from the 303 variables to make possible the exhaustive search algorithm , otherwise will be a big computation problem taking long time to make all the possible combinations. The initial dataset is reduced to a subset of 36 variables in case of transmission (Gearbox) that will be processed by 8 algorithms in parallel.

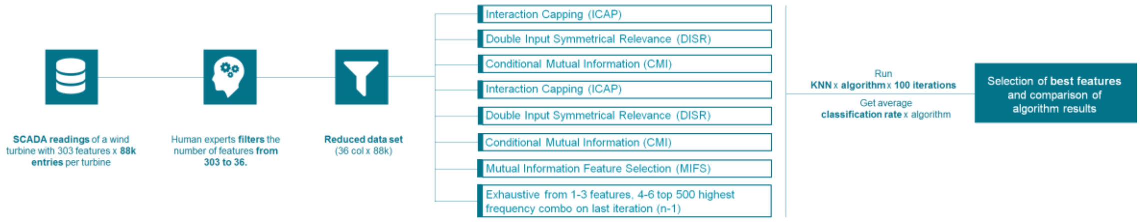

Figure 1: General Overview

<u>These algorithms are:</u>

- Conditional mutual information [1]
- Double input symmetrical relevance [2]
- Min-redundancy Max-relevance [3]
- Conditional mutual information maximization [4]
- Joint mutual information [5]
- Interaction Capping [6]
- Mutual information feature selection [7]
- Orthogonal Forward Regression [8]

The result of each algorithm is compared against the quasi-optimal method that is explained in following paragraph.

Quasi-optimal

The quasi-optimal algorithm is based on an exhaustive check of all possible combinations from one to three variables and evaluating the classification performance, only the one/two or three best variables are stored. From three to six variables the exhaustive method is change to a histogram based method.

This method consists on analyse the best 500 combinations results on the last iteration and reuse in combination with the new variable to be added in order to check its classification performance and select the best one. Again, the top 500 best results are stored for repeat the operation at the following iteration until six chosen variables. This method has been used since is almost impossible compute all the possible combinations and generate a model with almost any powerful desktop machine in a feasible time.

There are other experimental results with exhaustive method until four variables but the results were improved about 1-2% more than use Quasi-optimal.

Exhaustive + feature selection algorithm.

With the results of the eight feature selection algorithms, a hybrid approach was made to determine how good will be an exhaustive until three variables and then use a feature selection algorithm, in this case conditional mutual information that was the best performer overall. To develop this experiment the code of the library were modified in order to select exhaustively until three variables and use it at starting point for select new one’s (the already selected are outside from the available set to be selected).

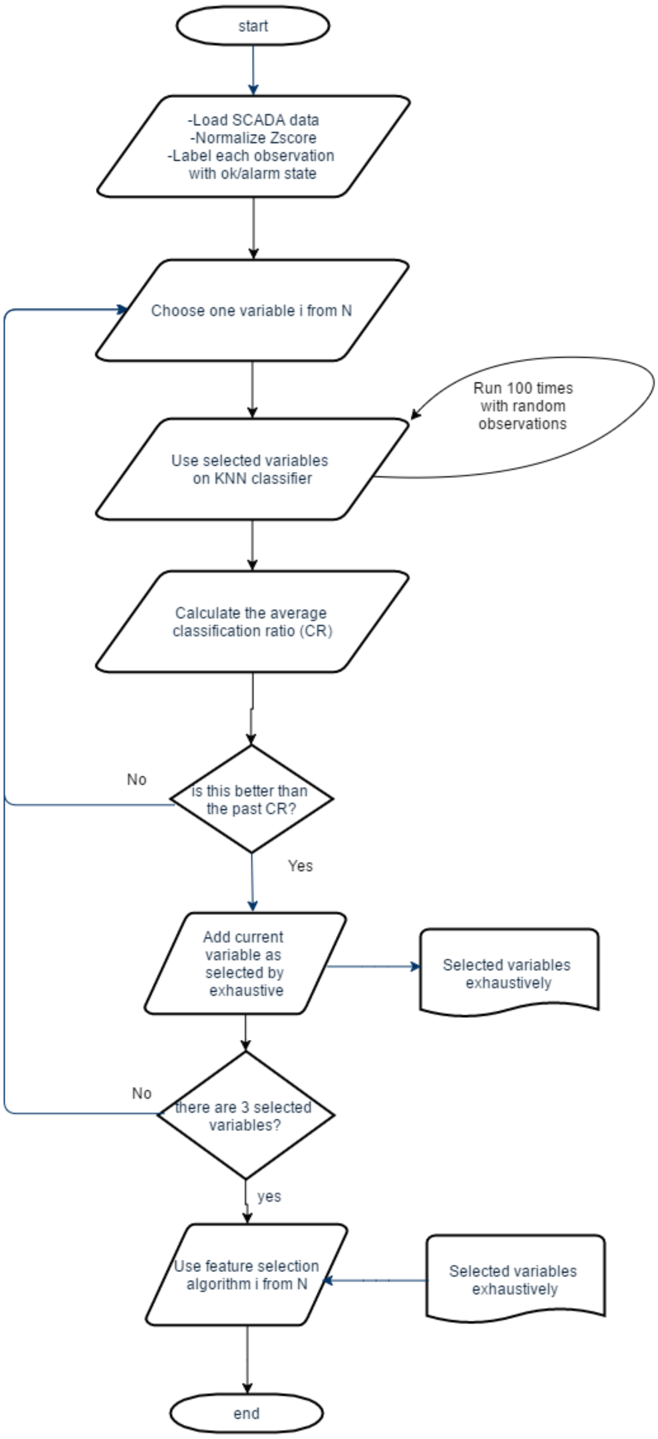

Figure 2: Exhaustive flowchart

KNN

In order to evaluate the feature selection impact in the classification results, a KNN [9] classifier was implemented in order to obtain the classification rate (CR) of the algorithm under test. The classification rate is measured after run KNN 100 times and the calculation of its average. The input observations are chosen randomly at each round from the main dataset, in order to avoid *over fitting* of the generated KNN model. KNN is basically a classification algorithm that assigns a label to each observation based on the neighbourhood labels, checking the distance and the frequency of the different labels that are near.

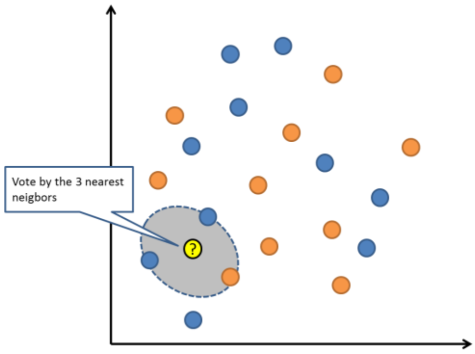

Figure 3: KNN with k=3 neighbor's example.

Results

Best feature selection algorithm.

This methodology has been applied to five Fuhländer wind turbines, for data gathered in 2014. A total of 8 algorithms were tested, this graphics shows the two betters versus quasi-optimal exhaustive method for the five wind turbines and for 1 to 6 features to be selected.

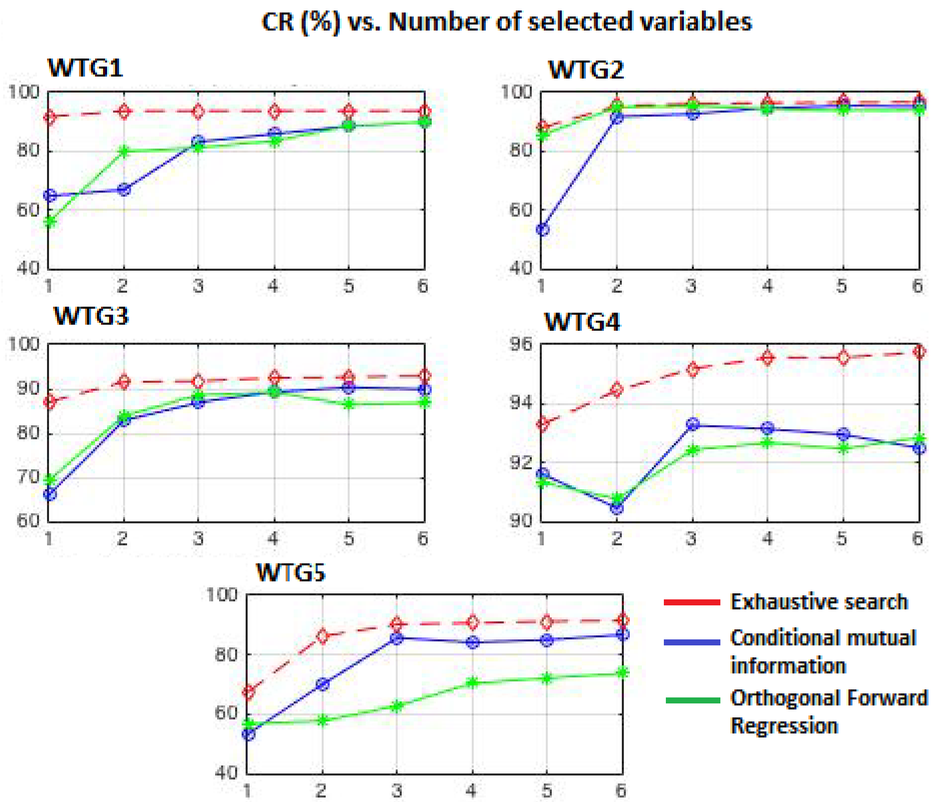

Figure 4: Best methods, classification results

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 and less computation bound) can help us to choose the best features, even from an bigger initial set of variables (the full 303 variables for example).

Hybrid approach combining the best feature selection algorithm

The following figures you may find the results of the best feature selection algorithm in this case, conditional mutual information (condMi) versus the hybrid approach starting from 3 exhaustive variables.

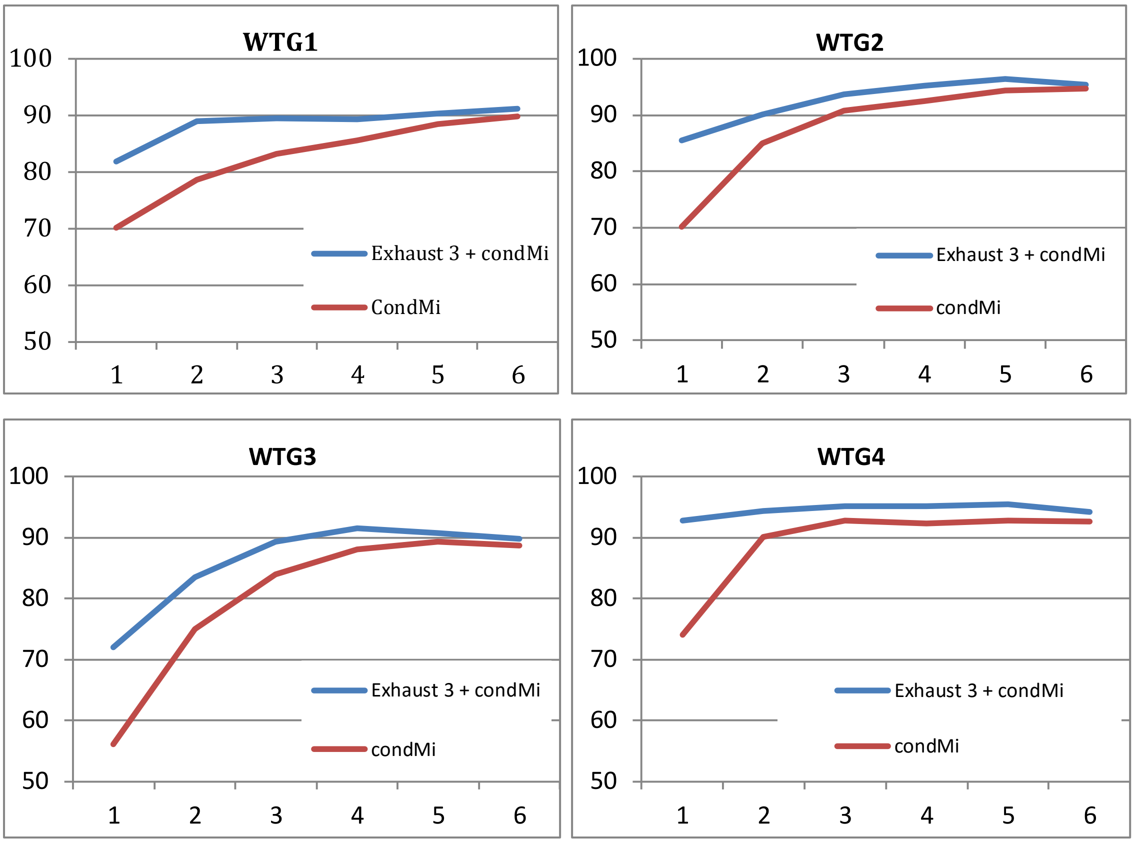

Figure 5 : WindTurbine 1 to 4 Hybrid vs Cond. MI

As shown, both approaches will deliver almost the same results once it arrives to six variables, the benefits of use exhaustive method is obtain better results when the input variable size is small, but the drawback is about the computation time needed to make all possible combinations until three, that is huge in this case, the order of 10 times the time of compute condMi.

3D representation to validate the selection

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.

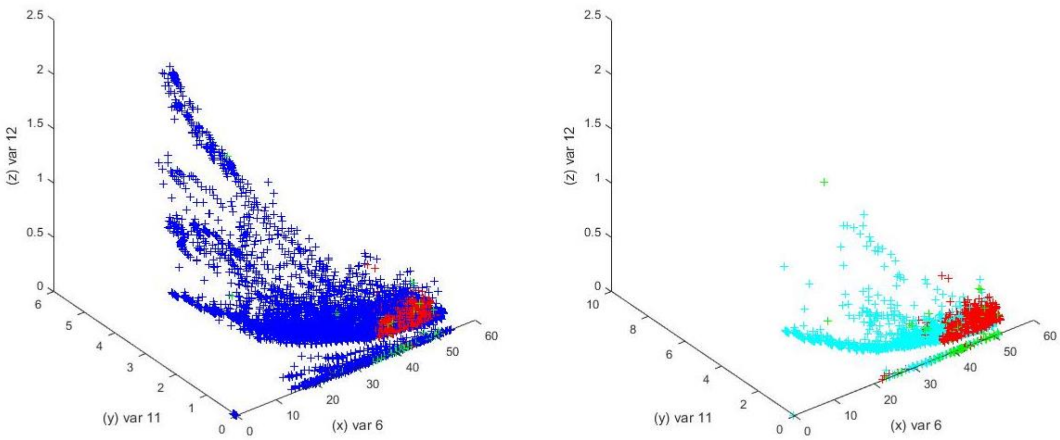

Figure 6: Variable values behaviour in space before failure.

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 evaluated versus an exhaustive method.

The results reveals that the variables have enough information combined that makes possible a pre-selection based of mutual information between a subset of them defining a specifically alarm as target. Also, there are evidences that for a large number of inputs, in this case six or more, search exhaustively all the possible combinations of variables versus use a suboptimal solution based on feature selection algorithms will deliver almost the same performance with less computation time.

References

1. H. Cheng, Z. Qin, C. Feng, Y. Wang, F. Li, "Conditional mutual information-based feature selection analyzing for synergy and redundancy", *ETRI Journal 33 (2011) 210–218*.
2. P. E. Meyer, G. Bontempi, "On the use of variable complementarity for feature selection in cancer classication", *Applications of Evolutionary Computing, Springer, 2006, pp. 91-102*
3. H. Peng, F. Long, C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy, Pattern Analysis and Machine Intelligence", *IEEE Transactions on 27 (2005) 1226–1238.*
4. F. Fleuret, "Fast binary feature selection with conditional mutual information", *The Journal of Machine Learning Research 5 (2004) 1531-1555*.
5. H. H. Yang, J. E. Moody, "Data visualization and feature selection: New algorithms for nongaussian data.", *NIPS, volume 99, Citeseer, pp.687–693.*
6. A. Jakulin, "Machine learning based on attribute interactions", *Ph.D. thesis, Univerza v Ljubljani, 2005*
7. R. Battiti, "Using mutual information for selecting features in supervised neural net learning", Neural Networks, *IEEE Transactions on 5 (1994) 537-550.*
8. Chen S, Billings S A and Luo W"Orthogonal least squares methods and their application to non-linear system identification*" Int. J. Control 50 1873–96 (*1989)
9. Dasarathy, B. V., "Nearest Neighbor (NN) Norms,NN Pattern Classification Techniques". *IEEE Computer Society Press*, 1990.