Transactions on Information and Communications Technologies vol 29, © 2003 WIT Press, www.witpress.com, ISSN 1743-3517

Identification of blade vibration causes in wind turbine generators

P. R. J. Campbell & K. Adamson
School of Computing and Mathematics, University of Ulster, UK

Abstract

For many years Data Mining techniques have been used to identify patterns and hidden information in datasets from a variety of domains. This paper will outline the initial application of DM techniques on a dataset from the Renewable Energy domain. For many years company's which operate wind farm sites have recorded and archived a range of data from numerous sources across wind farm sites and the wind turbine generators (WTG's) which make up these sites. The study presented in this paper encompasses a first phase of Knowledge Discovery from the dataset, making use of the complementary techniques Neural Networks (NN) and Rule Induction (RI). We discuss the challenges presented by the dataset in terms of selection and preparation of data and also introduce the format and meaning of data encompassed by the area of study, namely effects of blade vibration. Investigation of blade vibration has been used as an initial test of KD within the dataset as we are aware that blade vibration typically occurs within a specific range of wind speeds. By applying the NN and RI techniques we are able to support the theory that vibration occurs within a given range but we have also identified that other variables, namely rotor speed can contribute to blade vibration at lower than expected speeds.

1 Introduction

Knowledge Discovery has enjoyed significant success in the widest possible range of domains during the last decade, from telecommunications to supermarkets to medicine. Various terms have been used to describe the task of finding knowledge from dataset, including Data Mining and Knowledge Discovery. However, it is now generally accepted that all of these terms can be grouped under the more comprehensive title of KDD, Knowledge Discovery in

Transactions on Information and Communications Technologies vol 29, © 2003 WIT Press, www.witpress.com, ISSN 1743-3517

Databases, Fayyad *et al* [1]. Despite the wide ranging application of KDD amongst both industrial and academic areas, there still remain significant opportunities to spread the use and benefits of KDD to other areas, which provide interesting domain specific challenges.

Wind Energy companies have made use of advanced Information Systems for the retrieval and management of data, mainly in terms of storage and report production. However, little attention has been paid to the possibilities which KDD techniques can offer in identifying and modelling patterns which occur during site operation.

This paper looks at the application of KDD techniques to the task of analysing occurrences of blade vibration on a wind farm site in Ireland. The challenges presented by this study were twofold; firstly, apply KDD techniques to the Wind Energy domain, an extremely new and untouched area. Secondly, to identify the contributing factors which either individually or in combination cause blade vibration.

This work forms the first stages of ongoing research collaboration, between one of the United Kingdom and Ireland's leading Wind Energy Companies, B9 Energy Ltd and the University of Ulster.

The commercial data-mining package SPSS Clementine was used in order to speed development and facilitate experimentation within a Cross Industry Platform for Data Mining (CRISP-DM) methodology, as described by Chapman *et al* [2]. This enabled the modelling described in this paper to be developed during a truncated time frame, in order to prove the validity of KDD application to the domain.

In this paper the Neural Network and Rule Induction techniques are used to analyze occurrences of blade vibration, on a wind farm site in County Donegal, Republic of Ireland. The data is collected locally at the site and then downloaded to a central repository at B9 Energy Headquarters. The stages of data selection, coding and cleaning are described together with the interpretation of the results.

2 Task description

Blade vibration occurs naturally during periods of high wind speeds. However, it can be a significant risk to not only the WTG's blades but the entire structure Freebury and Musial [3] and Tangler [4]. At speeds of 14-17 m/s blades will begin to vibrate, passing these vibrations onto the nacelle and ultimately the tower.

SCADA (Supervisory Control and Data Acquisition) systems have preset boundaries for vibrations; a "blade vibration warning" is defined as a recorded reading of under 1.8 Hertz (oscillations per second) while a measurement of over 1.8 Hertz is recorded as a "blade vibration error". When the error event is registered, the WTG will initiate a timeout, applying brakes and stopping to allow the vibration to cease. Each WTG keeps a record of these errors and also holds a running count of vibration errors. Should ten errors be registered then the WTG automatically invokes a shutdown and requires manual intervention by site engineers.

While it is known that blade vibration occurs in high wind speeds, typically over 14m/s, this is not always the case. Indeed the peak production range for a typical WTG is 15-20 m/s of wind speed. We aim by conducting this investigation into the area of blade vibration to enhance understanding of its occurrence and perhaps also identify other contributing factors.

3 Data understanding

The study focused on one 25 WTG site in County Donegal. As it was an initial trial application of KD to the dataset, a small fraction of the dataset was selected, a single month (February 2001). However this still resulted in a subset with more than 100,000 records, in addition to the error records for the corresponding period.

3.1 Ten minute data

Intuitively this data is recorded from each WTG on a 10 minute basis. It records a wide range of information including:

- WTG ID
- Average Wind Speed
- Min & Max Wind Speed
- Average Power Output
- Average Rotor Speed
- Min & Max Rotor Speed

4 Data preparation

4.1 Conversion of data

As the dataset has been recorded directly by the SCADA system at the site it was held in a format dependant on SCADA type. While this is obviously useful in terms of site operation it means that the data is unusable in its current format within the chosen mining application (SPSS Clementine). As a result the first task was to convert the raw data into a more useable format. Having converted the data to standard MS Access table format it was decided that this approach while successful was too expensive in terms of processing; with numerous joins required to select the required data fields from numerous files. We therefore decided that the creation of two "flat files" from the relevant tables presented the most suitable and efficient option, these files could then be read directly into the mining engine.

4.2 Time series data

Following the creation of the two base files (error_log and turbine_log) it was necessary to remove the problems presented by time series data. Data held in these files was tagged with a TimeDate field which was unusable in the mining engine, which makes use of separate Time and Data fields. As a result, string concatenation was required to separate this single time series field into the required individual fields.

We made use of inbuilt derived nodes which enabled the splitting of the TimeDate field, in both files. The first stage was to spilt the string and discard the seconds portion leaving hours and minutes, the minutes were subsequently rolled back to the nearest 10 minute interval (to match 10 minute data from the turbine_log). Finally in this first stage the derived fields for hours and minutes were joined creating the new Time field. The remaining string values, the date in standard UK format of dd/mm/yyyy, were assigned to the new field of Date. Following the successful formatting of the two files they were merged, using 'turbine', 'date', and 'time' as merge fields. An example of the merged file is shown in Table 1.

Table 1: An example of the merge file created.

ErrorCode	Turbine #	WindSpeed	PowerOuput	RotorSpeed	Date	Time
Undef	8	12.59	574662.0	27.0	17/02/2001	06:00
Undef	8	11.98	528230.0	27.0	17/02/2001	06:10
Undef	8	12.03	525772.0	27.0	17/02/2001	06:20
Undef	8	11.07	462567.0	27.0	17/02/2001	06:30
Undef	8	11.69	504892.0	27.0	17/02/2001	06:40
Undef	8	11.79	514521.0	27.0	17/02/2001	06:50
Undef	8	11.85	526635.0	27.0	17/02/2001	07:00
Undef	8	12.32	550120.0	27.0	17/02/2001	07:10
Undef	8	11.78	510003.0	27.0	17/02/2001	07:20
Undef	8	11.54	496740.0	27.0	17/02/2001	07:30

4.3 Missing data

It is not uncommon for the encoded data to have fields that contain unknown or missing values. There are a variety of legitimate reasons why this can happen. And the filling of gaps in a data set is the focus of ongoing research. In this study occurrences may have been due to faulty monitoring equipment of an Individual WTG. However we were faced with the problem of how these missing values should be treated? There are a number of methods, as described by Goonatilake and Khebbal [5], for treating records that contain missing values:

1. Omit the incorrect field(s)
2. Omit the entire record that contains the incorrect field(s)
3. Automatically enter/correct the data with default values e.g. select the mean from the range
4. Derive a model to enter/correct the data
5. Replace all values with a global constant

A concurrent study is investigating the most accurate way in which to derive these missing values, however, after careful consideration it was decided that these generation algorithms were not, as yet sufficiently accurate at filling the missing values and as a result the best option initially was to ignore any records which contained missing values.

4.4 Creation of time related variables

Each record in the site_log (the result of merging the error and turbine logs) presents a 'snapshot' of the data at a given moment in time. Obviously it is then useful to examine not merely the values which immediately precede an error event, but also the behaviour of these values over a period of time. As a result time related variables were created; moving averages and rates of change for the three measurements contained by the log; wind speed, rotor speed and power output. Averages were calculated for these values over 3, 5 and 7 time intervals (30, 50 and 70 minute intervals). From these averages it was then possible to derive the rate of change for each variable.

5 Modelling techniques

The actions of human beings are governed by a combination of genetic information and learned knowledge. Evolution has imprinted information in our genes which aid our survival, while in some situations we use the information we have learned to reach decisions. Through this combination of methods, the human information processing process can be classed as a hybrid, in terms of AI classification Indurkhya and Weiss [6].

The ability to provide users with explanations of the reasoning process is an important consideration. Explanation facilities are essential for both user acceptance of the knowledge generated and also for the purpose of understanding whether the reasoning method is sound. As a result it was necessary to used two complementary modelling techniques; Neural Networks and Rule Induction.

5.1 Balancing records

A distribution table can be used to illustrate the occurrence of an event within a given dataset. Table 2 shows the occurrences of the 'blade vibration' event within the data subset under investigation. As can be seen in this case the proportion of occurrences compared to non-occurrences is 98.35% to 1.65%

Table 2: Distribution of occurrences.

Value	%	Occurrences
F	98.35	70284
T	1.65	1178

Transactions on Information and Communications Technologies vol 29, © 2003 WIT Press, www.witpress.com, ISSN 1743-3517

Obviously in such a case the application of a Neural Network (NN) would result in poor performance as NN's would only learn the non-event (error) case and ignore the event occurrence, when applied to such a biased dataset, Haykin [7]. A common solution to overcome this problem is the balancing of the data. As a result the balance was corrected, using all event-occurrences, but only 0.03% of the non-event-occurrences, Bishop [8].

Following the balancing of the data, the resulting file, containing the pre-processed (the derived, cleaned and balanced data subset) was now ready for use in the actual mining phase of the study.

5.2 Neural network testing

Using the Clementine package a Neural Network was created to analyse the data. Making use of the software's 'train net' node, Integral Solutions [9], the data was fed into a simple NN at the input layer, with the 'BladeVibrationError' field of the file set as 'out'. The remaining fields of the data are used as inputs in the learning process. Following the construction of the model the results can be easily analysed. As Figure 1 illustrates, the results from the simple NN show that the model was 96.77% accurate in predicting the occurrence of the 'blade vibration error'.

Results for output field BladeVibrationError
Comparing $N-BladeVibrationError with BladeVibrationError

Correct	*: 30*	*(96.77%)*
Wrong	*: 1*	*(3.23%)*
Total	*: 31*	

Figure 1: Neural network output.

5.3 Rule induction

Despite the NN technique showing that it was 96.77% accurate at predicting the occurrence of the event, as previously stated, it is frequently regarded with suspicion by non-technical observers and domain experts due to its 'black box' operation. As a result Rule Induction (RI) is frequently used as a complementary technique. RI creates an easy to follow decision tree representing a 'rule', for classifying data into different outcomes. This technique is therefore especially helpful to support results from NN application, Quinlan [10].

The C5 algorithm was selected for application in the RI testing stage, as its functionality is inbuilt into the Clementine environment. The process was used in the same way as the NN technique but its reasoning can be easily followed and interpreted. Once created and analysed it was found that the RI approach yielded the same accuracy rating as the NN approach, 96.77%, the simple rule is shown in Figure 2.

Transactions on Information and Communications Technologies vol 29, © 2003 WIT Press, www.witpress.com, ISSN 1743-3517

Interpretation of the results using the RI approach is much more intuitive and therefore more readily accepted by domain workers. Figure 2 supports the findings that if the wind speed (WindSpeed5 represents a moving average of wind speed, that is, the average value for wind speed over the last 5 time intervals, or 50 minute period) is greater than 15.536 m/s, then there is a 96.77% probability that blade vibration will occur.

```
Rules for T:
  Rule #1 for T:
    if WindSpeed5  >     15.536
         then  → T

Rules for F:
  Rule #1 for F:
    if WindSpeed5  =<    15.536
           then → F
```

Figure 2: Simple vibration rule.

5.4 Analysis of initial results

The results from both the NN and RI techniques show that for this site during the study period there is a strong pattern evident in the dataset, relating to blade vibration. These results can be further highlighted through the use of a histogram, as shown in Figure 3.

As with the RI technique, 'T' values represent an event occurrence while 'F' values show a non-occurrence of the event. The Y-axis shows the number of 'T' and 'F' values contained in the dataset, while the X-axis represents the wind speed (WindSpeed5) recorded. As can be seen there is a clear separation in event and non-event occurrences, with an apparent boundary at 15m/s. These results confirm common domain knowledge that indeed blade vibration does occur during sustained periods of high wind velocity. This confirmation validated the view of experts who at the outset of the study had identified that these were the factors which contributed to blade vibration. Secondly these results also proved that data mining did have a legitimate role to play in performance monitoring of WTG's.

However the approach highlighted a single outliner case which did not conform to the initial rationale for blade vibration. An occurrence of blade vibration was discovered during a period at which the wind's velocity was less than 10m/s, significantly outside the boundary point and also under the level identified by domain experts.

5.5 Outlier analysis

Having identified this outlier it was felt that its discovery warranted further investigation in order to fully explore the usefulness of data mining within the

Having identified this outlier it was felt that its discovery warranted further investigation in order to fully explore the usefulness of data mining within the wind energy domain. As a result the RI technique was reapplied to further investigate this new case.

As only one occurrence of this event had been identified it was necessary to again address the issue of balancing. However of this occasion the approach chosen was as follows all non-occurrences were used and 30 creations of the event occurrence were also used in the process, Compton *et al* [11]. Using the new dataset a further rule was induced, again using the C5 algorithm. The decision tree produced can be seen in Figure 4.

From the decision tree it is possible to infer that, if the rotor speed is greater than 18 rpm (revolutions per minute) and that average rotor speed over the last 30 minutes (RS3change, Rotor speed over the last 3 time intervals) is reducing, then blade vibration will occur.

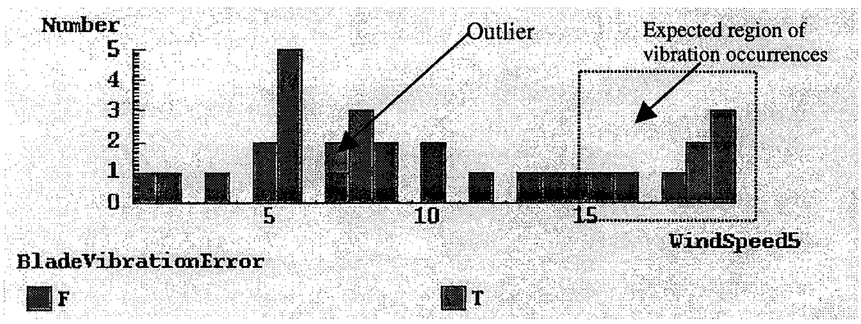

Figure 3: Histogram of vibration occurrences.

Rules for T:
Rule #1 for T:
if RotorSpeed > 18
and RS3change <= -2.807
then → T

Rules for F:
Rule #1 for F:
if RS3change > -2.087
then → F

Rule #2 for F:
if RotorSpeed <= 18
then → F

Default : → T

Figure 4: Outlier decision tree produced.

6 Conclusion

The application of modelling techniques to the dataset from this particular wind farm has identified two patterns:

- If the wind speed over the last 50 minutes is greater than or equal to 15.536 m/s, blade vibration occurs.

- At wind speeds less than 10 m/s, if the rate of change in the rotor speed is reducing over the last 30 minutes then blade vibration occurs.

The results of this initial study have yielded two results. The first has validated the existing knowledge of domain experts; this supports the methods applied in examining the dataset, and also provides evidence to support the knowledge of the domain experts which has been gained over many years of supporting wind farm sites.

The second result has identified a correlation between wind speed and rotor speed. This has not previously been identified by domain experts, and is currently under investigation to ascertain, whether the knowledge discovered during this study can be used to provide indicators of component degradation or failure.

The study has also clearly proven that Data Mining and Knowledge can yield results when applied to datasets from the wind energy industry. We believe that the expected and unexpected results from this study provide conclusive proof that Knowledge Discovery should continue to be applied to datasets from this area.

7 Further work

This study has examined only one type of error and included only three variables; wind speed, rotor speed and power output. There are many other faults which can be encountered on WTG's and numerous factors which can determine or influence their occurrence. Obviously further study of these events and indeed further study on a wider scale and over a longer time period, into blade vibration is required. Immediately three areas of interest have been identified for further study:

- Blade Vibration Warning – the event which occur before Blade vibration
- Effects of minor faults – do any minor faults contribute to a more serious event, and if so can a pattern be identified which could be used for the purposes of predictive maintenance.
- The consequences of manual intervention – does a pattern exist in the data, which is directly related to the manual intervention of engineers, for example, following the stopping of a WTG for servicing.

Acknowledgements

The authors wish to thank B9 Energy Ltd for its continued assistance and invaluable support and expertise. Many thanks also to Shirley Carolan, for her hard work during the investigation.

References

[1] Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., Uthurusamy, R., (Editors), "*Advances in Knowledge Discovery and Data Mining*", AAAI/MIT Press, 1996.

[2] Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., "*CRISP-DM 1.0 Step-by-step data mining guide*", USA, SPSS Inc, 2000.

[3] Freebury, G. E.; Musial, W. "*Determining Equivalent Damage Loading for Full-Scale Wind Turbine Blade Fatigue Tests*". Collection of the 2000 ASME Wind Energy Symposium Technical Papers Presented at the 38th AIAA Aerospace Sciences Meeting and Exhibit, 10-13 January 2000, Reno, Nevada. New York: American Institute of Aeronautics and Astronautics, Inc. (AIAA) and American Society of Mechanical Engineers (ASME); pp. 287-296; NREL Report No. 32247, 2000.

[4] Tangler, J. L., "*Evolution of Rotor and Blade Design*", NREL Report No. CP-500-28410, 2000.

[5] Goonatilake, S., Khebbal, S, "*Intelligent Hybrid Systems*", John Wiley and Sons, 1995.

[6] Indurkhya, N., Weiss, S.M., "*Predictive data mining: a practical guide*", Moragn Kaufmann Publishers Inc, 1998.

[7] Haykin, S., "*Neural Networks: a comprehensive foundation*", New York Macmillan, 1994.

[8] Bishop, C.M., "*Neural Networks for pattern recognition*", Oxford University Press, 1996.

[9] "*CLEMENTINE User Guide*", Integral Solutions Ltd, England, 1996

[10] Quinlan, J.R., "*C4.5: Programs for machine learning*", Morgan Kaufmann Publishers, 1993.

[11] Compton, P.J., Horn, K.A. and Lazarus, L., Quinlan, J.R., "*Inductive knowledge acquisition: A case study*", In J.R. Quinlan (Editor), Applications of Expert Systems, Addison-Wesley, UK, 1987.