An evaluation of different data mining methods for forecasting wind farm power

Gerry Wiener, J. M. Pearson, B. Lambi, W. Myers and K. Goodrich
National Center for Atmospheric Research, Boulder, CO

In forecasting wind farm power output, it is important to obtain an accurate farm power output estimate based on given forecast winds. Generally, the manufacturer’s turbine power curves are applied to obtain this estimate especially in cases when observed wind data at farms are not available. In this paper we will compare manufacturer power curve performance against the performance of a number of different data mining techniques including regression trees, KNN nearest neighbor and random forest regression based on using actual observed wind and power data. The modeling applied here differs from most traditional power curve applications since previous wind and power information are both utilized in order to forecast future power. Mean absolute error for different techniques will be presented and the application of these techniques for forecasting winds and power will be discussed.

Overview

NCAR Wind/Power System
- Utilizes forecast winds from multiple numerical weather models
- WRF, GFS, NAM, RUC, GEM, MMS
- Applies statistical dynamic MOS technology DiCast® to formulate a tuned wind forecast using the model forecast winds at each wind turbine
- Applies statistical model to evaluate power at each wind turbine
- Sums turbine powers to yield farm, connection node and regional power forecasts

Power Curve Estimation Approach
The simplest approach is to apply the manufacturer’s power curve for a given turbine to a particular forecast wind to estimate turbine power
Node however that observed wind turbine power output does not in practice resemble the idealized manufacturer’s power curve.
(See Figures 1 and 2) Power curves representing the total power at a farm connection node tend to be better behaved owing to error cancellation but there are still problems. (See Figure 3)

Table Description
The tables present mean absolute error (MAE) in kilowatts for the turbine study and megawatts for the connection node study.

The Different Approximation Techniques
- Curve represents results derived from using the standard power curve in going from wind to power.
- Tree represents results derived from using a regression tree. The training variables used were previous wind, previous power, observed wind and the target variable was observed power.
- Forest represents results derived from using a regression forest with 200 trees. The training and target variables utilized are identical to those of the regression tree application.
- KNN represents results derived from using KNN nearest neighbor with k=25. Again the training and target variables utilized are identical to those of the regression tree application.

Results

<table>
<thead>
<tr>
<th>Turbine Power MAE in KW</th>
<th>Curve</th>
<th>Tree</th>
<th>Forest</th>
<th>KNN</th>
<th>Persistence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turbine 1</td>
<td>64.4</td>
<td>17.6</td>
<td>10.2</td>
<td>95.4</td>
<td>78.3</td>
</tr>
<tr>
<td>Turbine 2</td>
<td>48.4</td>
<td>16.7</td>
<td>16.3</td>
<td>83.2</td>
<td>74</td>
</tr>
<tr>
<td>Turbine 3</td>
<td>35.2</td>
<td>17.2</td>
<td>16.7</td>
<td>93.1</td>
<td>73.8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Turbine Power Normalized MAE</th>
<th>Curve</th>
<th>Tree</th>
<th>Forest</th>
<th>KNN</th>
<th>Persistence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turbine 1</td>
<td>0.020</td>
<td>0.012</td>
<td>0.013</td>
<td>0.060</td>
<td>0.052</td>
</tr>
<tr>
<td>Turbine 2</td>
<td>0.022</td>
<td>0.012</td>
<td>0.012</td>
<td>0.055</td>
<td>0.049</td>
</tr>
<tr>
<td>Turbine 3</td>
<td>0.023</td>
<td>0.011</td>
<td>0.012</td>
<td>0.062</td>
<td>0.049</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Connection Node Power MAE in MW at a 102 MW Farm</th>
<th>Curve</th>
<th>Tree</th>
<th>Forest</th>
<th>KNN</th>
<th>Persistence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Connection Node Power MAE</td>
<td>2.59</td>
<td>0.695</td>
<td>1.28</td>
<td>3.71</td>
<td>3.7</td>
</tr>
</tbody>
</table>

Power Curve Approximation Techniques

The power curves depicted in the previous slide demonstrate that estimating power using wind speed alone can lead to significant errors.

Using Wind and Power to Forecast Power
Owing to the inaccuracy mentioned above, we decided to investigate incorporating previous wind speed and power data in addition to current wind speed data in order to reduce wind to power conversion error.

Techniques Explained:
- Power Curve
- Regression Tree (Cubist)
- Random Forest Regression (R package)
- KNN Nearest Neighbor
- Persistence

Summary of Findings

Turbine Results
The turbine results illustrate that the regression tree and regression forest produce the lowest errors for the set of techniques chosen here and for the data set under consideration. Applying the industrial power curve directly to the turbine wind increased the error by a factor of approximately 2 to 3 times. The KNn algorithm was in line with persistence and trailed in performance. Note that the Cubist implementation of the regression tree was significantly more efficient than the R implementation of random forest outperforming it by more than a factor of 100.

Connection Node Results
The connection node results illustrate again that both the regression tree and regression forest are yielding the best error results out of the techniques evaluated. Note that the tuning parameters for the random forest were not evaluated exhaustively so there is potential for improved performance. Still the application of random forest generally requires some additional tuning and significantly longer run times.