Renewable Energy 35 (2010) 2362–2369

Contents lists available at ScienceDirect

Renewable Energy

journal homepage: www.elsevier.com/locate/renene

Short-term prediction of wind power with a clustering approach

Andrew Kusiak*, Wenyan Li

Department of Mechanical and Industrial Engineering, 3131 Seamans Center, The University of Iowa, Iowa City, IA 52242–1527, USA

ARTICLE INFO

Article history:
Received 30 October 2009
Accepted 23 March 2010
Available online 18 April 2010

Keywords:
Wind turbine
Wind energy
Data-mining
Clustering
Power prediction
Parameter selection

ABSTRACT

A clustering approach is presented for short-term prediction of power produced by a wind turbine at low wind speeds. Increased prediction accuracy of wind power to be produced at future time periods is often bounded by the prediction model complexity and computational time involved. In this paper, a trade-off between the two conflicting objectives is addressed. First, a set of the most relevant parameters (predictors) is selected using the underlying physics and pattern immersed in data. Five scenarios of the input space are created with the *k*-means clustering algorithm. The most promising clustering scenario is applied to produce a model for each clustered subspace. Computational results are compared and the benefits of cluster–specific (customized) models are discussed. The results show that the prediction accuracy is improved the input space is clustered and customized prediction models are developed.

1. Introduction

Knowing the power to be produced by a wind turbine at future time horizons is of interest to the rapidly expanding wind industry [1]. Wind power forecasts are used as input for various tools, e.g., management of power dispatch and control of wind turbines [2,3].

The state-of-the-art approaches to wind power forecasting have been published in [4] with more recent updates included in [5]. Models used for forecasting wind power are categorized as physics-based models, statistical models, and spatial correlation models [3,6–10]. Data-mining algorithms offer a promise to conquer the unresolved gap of handling the dynamic nature of wind [11].

The published literature on data-mining in wind power is growing, with Neural Networks (NNs) becoming the widely used algorithm. NN models can be used to estimate power output as a function of wind turbine parameters (e.g., wind speed, generator torque) and time delay of the corresponding parameters (e.g., power itself, wind speed) [12,13]. Wind speed, relative humidity, and time were used as input variables to train an NN model in power prediction applications [14,15]. The recurrent multilayer-perceptron NN was applied for power prediction in [16]. Long- and short-term prediction of power using the *k*-nearest neighbor (*k*-NN) algorithm was presented in [11,12]. Analysis and estimation of power based on cluster analysis was reported in [17].

This paper is organized in eight sections. Section 2 describes the data used for this paper. The candidate parameters of interests to the studied domain are selected using known equations. Section 3 discusses the proposed clustering approach (model customization) for short-term power prediction. Parameter selection by data-mining algorithms is presented in Section 4. Section 5 discusses and compares five scenarios of clustering the input space. Section 6 presents models extracted from data of each clustered subspace. Comparison between the proposed model and the model extracted in typical ways is discussed in Section 7. Section 8 concludes the paper.

2. Data description and parameter selection

Wind turbine data is usually collected by a Supervisory Control and Data Acquisition (SCADA) system. Though the data sampling frequency may be relatively high (e.g., 20 Hz), the data is averaged into time intervals, e.g., 10 s, 30 s, or 10 min, that are suitable for various applications. The data used in this paper was collected at 10 s intervals (called 10 s data) at a 1.5 MW wind turbine (randomly selected) for a period of seven days. For the selected wind turbine, the cut-in speed is 3.5 m/s, the rated speed is 12.5 m/s, and cut-out speed is 21 m/s. From the view of turbine operations, wind speed in the range [3.5 m/s, 12.5 m/s] is of interest to industry. Thus, the data with a wind speed lower than 3.5 m/s or higher than 12.5 m/s have been excluded from analysis in the research reported in this paper. Data points with a negative power output have been also deleted. The data from the first five days (approximately 2/3 of all data) was used to extract models, and the data from the remaining two days (approximately 1/3 of all data) was used for test and validation models. The data set used in this research is characterized in Table 1.

* Corresponding author.
E-mail address: andrew-kusiak@uiowa.edu (A. Kusiak).

0960-1481/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.renene.2010.03.027

Table 1
Description of the training and test data.

Data set	Start time	End time	Time interval	Number of data points
Training	8/8/07 12:00 AM	8/12/07 12:00 AM	10 s	30 354
Test	8/13/07 12:00 AM	8/15/07 12:00 AM	10 s	15 860

The data available for this research included numerous parameters of a wind turbine. Some of these parameters could have a potential impact on the prediction accuracy of wind power to be generated at 10 s intervals.

Based on the basic wind power equation [18], five parameters are selected as candidates, including wind speed v, blade pitch angle β, generator torque T, and rotor speed ω_r. Considering the fact that the system inertia could be significant, the power output P_a is also included. The air density ρ and rotor radius R are regarded here as constants. The initially selected parameters are listed in Table 2.

Since time delay is considered as having an impact on the model's accuracy, it also considered, and thus further parameter selection is accomplished with data-mining algorithms (see Section 4).

3. Proposed methodology

The input data (controllable, non-controllable, and performance parameters) representing the input space undergo parameter selection and clustering. Based on the data in each cluster, a model is produced with data-mining algorithms. The number of input parameters n (dimension) and the number of instances N define the input space IS_n. For each parameter in the input space, there are $N = 30\ 354$ training instances 1and $N = 15\ 860$ test instances (see Table 1). Each of the models 1,…,k predicts power output $PO(t)$ at time t.

The steps of the proposed methodology are discussed next.

3.1. Parameter selection

The five parameters listed in Table 2 partially describe the input space. To simplify the input space, the same number m of past states is considered for each of the five parameters. The input space IS_n for the m past states of five parameters listed in Table 2 is defined by vector (1).

$$IS_n = [v(t-1),\ldots,v(t-m),x_1(t),\ldots,x_1(t-m),x_2(t),\ldots,x_2(t-m),\quad y_1(t-1),\ldots,y_1(t-m),y_2(t-1),\ldots,y_2(t-m)] \quad (1)$$

Of all n predictors, the most significant are selected with a NN algorithm (see Section 4).

3.2. Clustering input space

3.2.1. Clustering training data set

In this section, the training data set is clustered into k subspaces. Based on parameters selected for clustering, the following five data processing scenarios are considered:

a) Clustering on wind speed estimated by its one past state
b) Clustering on wind speed estimated by the time series model
c) Clustering on generator torque $x_2(t)$
d) Clustering on generator torque $x_2(t)$ and rotor speed $y_2(t-1)$
e) Clustering on generator torque $x_2(t)$, rotor speed $y_2(t-1)$, and power output $y_1(t-1)$

Table 2
List of parameters selected for wind speed estimation.

Parameter type	Parameter name	Abbreviation	Symbol	Unit
Non-controllable	Wind speed	WS	v	m/s
Controllable	Blade pitch angle	BPA	x_1	°
	Generator torque	GT	x_2	Nm
Performance	Power output	PO	y_1	kW
	Rotor speed	RS	y_2	rpm

The first two scenarios explore the impact of wind speed on the accuracy of power output predictions at 10 s intervals. As the wind speed $v(t)$ at future time t is not known, two estimation methods have been applied. The first one (item (a) above) uses wind speed at one past state $v(t-1)$, and the second one is based on the time series model. Prediction of wind speed with the time series models has been proven to be accurate (see [19]).

The final three scenarios of clustering the input space originate in predictors' importance (see Section 4). The first three most significant parameter states determined by the data-mining algorithms are the generator torque $x_2(t)$, rotor speed $y_2(t-1)$, and power output $y_1(t-1)$ (see Section 4). All parameters are studied for impact on clustering the input space.

3.2.2. Clustering the test date set

For the $n$dimensional space, the center (centroid) of the ith cluster of the training data is denoted as $[x_1^i, x_2^i, \ldots, x_n^i]$, where i is the number of the cluster satisfying $1 \le i \le k$. To balance the bias due to the variability of the input data the values of $[x_1^i, x_2^i, \ldots, x_n^i]$ have been normalized in the interval [0,1].

Thus, the distance from a normalized instance $[z_{1,\ \mathrm{normalized}},\ldots, z_{n,\ \mathrm{normalized}}]$ to the ith cluster centroid of the training data is defined in (2).

$$D^i = \sqrt{\left(z_{1,\mathrm{normalized}} - x^i_{1,\mathrm{normalized}}\right)^2 +,\ldots,+ \left(z_{n,\mathrm{normalized}} - x^i_{1,\mathrm{normalized}}\right)^2} \quad (2)$$

The aim is to find, for each data instance, a cluster with the minimum distance between the instance and the cluster centroids. The clustering algorithm of the test data set is shown in Table 3.

As illustrated in Table 3, $D^{\min}$ is initially set to 1, and then it is replaced with a shorter distance found. In this way, each instance from the test data set is assigned to the closest cluster.

Table 3
Algorithm for clustering test instances.

```
Begin
For i = 1 to k
  D^min = 1
D^i = sqrt((z_1,normalized − x^i_1,normalized)^2 +,…,+(z_n,normalized − x^i_n,normalized)^2)
    If D^i ≤ D^min
      Let
D^min = D^i, [z_1,normalized,…,z_n,normalized] = [z^i_1,normalized,…,z^i_n,normalized]
    Else
Next i
End
```