Applied Energy 188 (2017) 56–70

Contents lists available at ScienceDirect

Applied Energy

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

Deep learning based ensemble approach for probabilistic wind power forecasting

Huai-zhi Wang [a], Gang-qiang Li [a], Gui-bing Wang [b,*], Jian-chun Peng [a], Hui Jiang [c], Yi-tao Liu [a]

[a] The College of Mechatronics and Control Engineering, Shenzhen University, Shenzhen 518060, China
[b] Shenzhen Key Laboratory of Urban Rail Transit, The College of Urban Rail Transit, Shenzhen University, Shenzhen 518060, China
[c] The College of Optoelectronic Engineering, Shenzhen University, Shenzhen 518060, China

HIGHLIGHTS

- Convolutional neural network is designed for probabilistic wind power forecasting.
- Ensemble technique is used to cancel out the diverse errors of point forecasters.
- The model misspecification and data noise in wind power are separately evaluated.
- The competitive performance and robustness of the proposed method were proved.

ARTICLE INFO

Article history:
Received 9 September 2016
Received in revised form 8 November 2016
Accepted 26 November 2016

Keywords:
Convolutional neural network
Ensemble
Probabilistic wind power forecast
Deep learning
Wavelet transform

ABSTRACT

Due to the economic and environmental benefits, wind power is becoming one of the more promising supplements for electric power generation. However, the uncertainty exhibited in wind power data is generally unacceptably large. Thus, the data should be accurately evaluated by operators to effectively mitigate the risks of wind power on power system operations. Recognizing this challenge, a novel deep learning based ensemble approach is proposed for probabilistic wind power forecasting. In this approach, an advanced point forecasting method is originally proposed based on wavelet transform and convolutional neural network. Wavelet transform is used to decompose the raw wind power data into different frequencies. The nonlinear features in each frequency that are used to improve the forecast accuracy are later effectively learned by the convolutional neural network. The uncertainties in wind power data, i.e., the model misspecification and data noise, are separately identified thereafter. Consequently, the probabilistic distribution of wind power data can be statistically formulated. The proposed ensemble approach has been extensively assessed using real wind farm data from China, and the results demonstrate that the uncertainties in wind power data can be better learned using the proposed approach and that a competitive performance is obtained.

1. Introduction

Due to the continuous decrease in the storage capacity of fossil fuel, the energy crisis is becoming more significant than ever [1]. Therefore, to mitigate the energy crisis, regulatory acts that encourage the use of renewable energy have been promoted worldwide. Among the renewable energy resources, wind energy, as an alternative to fossil energy, has attracted much attention due to its beneficial impacts on climate change mitigation and environmental pollution reduction [2]. Coupled with its mature technology, wind energy has experienced an unexpected annual growth on a global scale. Wind energy can be used to drive engines directly and provide rural energy services. In [3], a novel mean flow acoustic engine with a cross-junction configuration was designed to convert wind energy in a pipeline into acoustic energy, and its efficiency was numerically analyzed in [4] by using computational fluid dynamics method. In practice, wind energy is mainly utilized to mechanically power generators for electricity. The annual growth rate of worldwide wind power has been between 20% to 35% per year since 2000 [5]. However, due to the chaotic nature of the earth's atmosphere, wind generated power always exhibits nonlinear and non-stationary uncertainties, which pose great challenges for the management and operations of electric power and

* Corresponding author.
E-mail addresses: wanghz@szu.edu.cn (H.-z. Wang), ligq@szu.edu.cn (G.-q. Li), wanggb@szu.edu.cn (G.-b. Wang), jcpeng@szu.edu.cn (J.-c. Peng), huijiang@szu.edu.cn (H. Jiang).

http://dx.doi.org/10.1016/j.apenergy.2016.11.111
0306-2619/© 2016 Published by Elsevier Ltd.

Nomenclature

ACE	average coverage error	BP	back-propagation algorithm
CNN	convolutional neural network	CRPS	continuous ranking probability score
DBM	deep restricted Boltzmann machine	IS	interval sharpness
MWWF	milky way wind farm	NN	neural network
PI	prediction interval	PINC	prediction interval nominal confidence
QR	quantile regression	SAE	stacked auto-encoder
SIWF	Shangchuan island wind farm	SVM	support vector machine
WPF	wind power forecasting	WT	wavelet transform
A_n	wavelet approximation signal	CDF_i	cumulative distribution function at time step i
D_n	wavelet detail signal	DS_{du}	dataset used for data noise uncertainty evaluation
DS_{mu}	dataset used for model uncertainty evaluation	E_m	squared-error loss function considering m batches
GD	Gaussian distribution	$H(\cdot)$	indicative function
I_i^α	PI at time step i given PINC = $100(1-\alpha)\%$	L_h^α	lower bound of PI given target h and PINC
M_{du}	mean of data noise uncertainty	M_ε	mean of the uncertainty signal $\varepsilon(\boldsymbol{x}_i)$
N_E	number of ensembles	N_M	number of selected input maps
N_S	number of training samples	U_h^α	upper bound of PI given target h and PINC
$\boldsymbol{W}$	CNN's weight matrix	$\boldsymbol{W_{con}^l}$	weight matrix at lth convolution layer
$\boldsymbol{W_{log}^L}$	weight matrix at Lth logistic regression layer	WS_i^α	wind speed at time step i given PINC
T	length of the signal required to be decomposed	b_j^l	bias of jth output map at lth layer
$\boldsymbol{b}$	CNN's bias matrix	$\boldsymbol{b_{con}^l}$	bias matrix at lth convolution layer
$\boldsymbol{b_{log}^L}$	bias matrix at Lth logistic output layer	c_j^l	additive bias of jth output map at lth layer
$\boldsymbol{c}$	CNN's additive bias matrix	$\boldsymbol{c_{sub}^l}$	additive bias matrix at lth sub-sampling layer
d	output vector size of training samples	$down\ (\cdot)$	down-sampling function
$f(\cdot)$	output activation function	$g(\cdot)$	signal required to be decomposed by wavelet
h_j^i	the jth target in ith training sample	len	length of a given map
m	mini-batch size of training sample	r_i	indicator of prediction interval coverage probability
t	discrete time step	$\boldsymbol{u^L}$	output vector of the neurons in $(L-1)$ layer
$up\ (\bullet)$	up-sampling function	wid	width of a given map
$\boldsymbol{w}_{i,j}^l$	weight matrix at lth layer connecting the ith input map and jth output map	$\boldsymbol{x}_i^l$	the ith input map at lth layer
		$\boldsymbol{x}^{L-1}$	the output of the neurons at $(L-1)$th layer
$\boldsymbol{x}_{i,j,k}$	the ith input in jth input map at kth layer	$\boldsymbol{y}_{i,j,k}$	the ith output in jth output map at kth layer
$\widehat{y}(\boldsymbol{x_i})$	mean of the estimated model uncertainty	y_j^i	the jth output in ith training sample
$\boldsymbol{y}_j^l$	the jth output map at lth layer	$z_{1-\alpha/2}$	critical value of a Gaussian distribution function
$\widehat{y}(\boldsymbol{x_i})$	output of the jth deep CNN model	$\boldsymbol{\beta}$	CNN's multiplicative bias matrix
α	confidence level parameter	β_j^l	multiplicative bias of the jth output map at lth layer
$\boldsymbol{\beta_{sub}^l}$	multiplicative bias matrix at lth sub-sampling layer	δ_i^α	width of the PI at time step i given PINC
$\boldsymbol{\gamma_{len,wid}}$	average filter parameter matrix with size $len \times wid$	η	learning rate
$\varepsilon(\boldsymbol{x}_i)$	uncertainty given the input $\boldsymbol{x}_i$	υ	scaling variable
κ	translation variable	σ_{du}^2	variance of data noise uncertainty
$\phi(\bullet)$	mother wavelet function	σ_h^2	variance of total forecasting error
σ_{mu}^2	variance of model uncertainty		
σ_ε^2	variance of the uncertainty signal $\varepsilon(\boldsymbol{x}_i)$		

energy systems. It is demonstrated in [6] that the impact of these uncertainties on power system operations can be, to a certain degree, mitigated via advanced WPF methods, which are considered to be the most promising solutions for the integration of a large amount of wind energy into power grids. Aimed at this task, three typical methodologies for WPF have been proposed in the literature, including physical modeling, statistical methods and soft-computing techniques.

Physical modeling methods try to establish an accurate mathematical model for WPF using various geographical and meteorological information. However, this type of approach may not be applicable for practical real-time prediction tasks due to the high amount of calculation costs involved [7,8], whereas statistical approaches manage to develop an optimal relationship between future wind power and historical samples via error minimization. In [9], a generalized WPF model was proposed based on time-varying threshold autoregressive moving average, and the efficiency was numerically analyzed. In [10], a hybrid statistical approach in combination with empirical wavelet transform, partial auto-correlation function and Gaussian process regression was proposed. Simulation results indicate that the suggested approach, i.e., the generalized WPF model, performed the best among the three compared methods. In addition, soft-computing techniques, such as the artificial neural network [11,12] and Elman neural network [13], were utilized for WPF. In [14], a WPF model based on extreme machine learning was presented to evaluate wind power density. In [15], a multi-layer neural fuzzy network was mooted for hour-ahead WPF, and the model parameters were well-trained by using simultaneous perturbation stochastic approximation. In [16], a hybrid model based on wavelet packet technique and artificial neural network was originally proposed, and the model parameters were optimized by using crisscross optimization algorithm. In [17], reproducing kernel Hilbert space based probabilistic WPF method was proposed and the performance was evaluated by CRPS. In [18], the randomness and uncertainty of wind energy were quantitatively evaluated using Gaussian process regression and teaching learning optimization. In [19], a general framework based on k-nearest neighbors algorithm and kernel