Energy Conversion and Management 163 (2018) 384–406

Contents lists available at ScienceDirect

Energy Conversion and Management

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

Correlation aware multi-step ahead wind speed forecasting with heteroscedastic multi-kernel learning

Yun Wang[a], Zongxia Xie[b,*], Qinghua Hu[a], Shenghua Xiong[c]

[a] School of Computer Science and Technology, Tianjin University, Tianjin, China
[b] School of Computer Software, Tianjin University, Tianjin, China
[c] Department of Management and Economics, Tianjin University, Tianjin, China

ARTICLE INFO

Keywords:
Multi-step ahead wind speed forecasting
Heteroscedasticity
Task correlation
Multi-kernel learning
Matrix-variate Gaussian
Variational Bayesian

ABSTRACT

For the management of wind energy, wind speed forecasting is often required. Accurate multi-step ahead wind speed forecasts make the power system be adjusted timely and properly to ensure the stable and efficient operation of power system. Currently, various techniques have been developed for multi-step ahead wind speed forecasting. However, the correlation among different forecasting steps is often neglected in current multi-step ahead wind speed forecasting approaches, and the characteristic of heteroscedasticity in wind speed forecasting errors is usually not taken into consideration. In this work, a novel multi-step ahead forecasting method based on multi-kernel learning is developed. This method considers the task correlation, which represented by the covariance of multi-step ahead forecasting tasks, as well as the heteroscedasticity of forecasting errors. The optimization is solved within the framework of variational Bayesian. Thus, a correlation aware multi-step ahead wind speed forecasting technique with heteroscedastic multi-kernel learning is designed. In this paper, the experimental results in different wind farms and different seasons prove that the regression model considering the characteristics of multi-step ahead wind speed forecasting, task correlation and heteroscedasticity, will produce more accurate forecasts than the other models as for two to six-ahead wind speed forecasting. However, it is difficult to tell which characteristic is more important from the forecasting results. So, the regression model considering both of them will be more reasonable. Moreover, the training time of the proposed model is more than 10 min but less than 20 min. Thus, two to six-step ahead wind speed forecasts can be used in some practical applications, such as the load dispatch planning and the load increment/decrement decisions.

1. Introduction

Since the 21st century, demands for energy have rapidly increased with the development of economy, and energy industry is growing fast. Wind energy has extracted more and more attention in recent years. However, as the proportion of wind power in the whole electricity supply increases constantly, the corresponding drawbacks have gradually emerged. The random fluctuation and intermittence characters of nature wind result in the unstable wind power. Large wind disturbance will cause voltage and frequency fluctuations. More seriously, the whole power system will lose stability [1].

Wind power forecasting is the premise for the stable development of wind power industry. Accurate wind power forecasts will help adjust the power dispatching plans in time, which will reduce the adverse impact of wind power on the power grid and improve the operation benefit of wind farms in the meanwhile. Before getting the wind power forecasts, wind speed forecasts are required [2]. Lots of wind speed forecasting models have been designed, but no general wind speed forecasting model can be applied to all wind regimes. Generally, they can be grouped into three categories: physical models, statistical methods and hybrid models [3].

Physical models can infer the final forecasted wind speed by physical equations when given the physical information and the outputs of numerical weather prediction (NWP) models [4,5]. Unlike the physical models, statistical models only utilize the historical wind speed data. The typical statistical models contain autoregressive (AR) model [6], autoregressive moving average (ARMA) model [7] and autoregressive integrated moving average (ARIMA) model [8]. Some variants, such as fractional-ARIMA and Hammerstein autoregressive, are also proposed to forecast wind speed [9,10].

The conventional statistical approaches just show the linear relationship between the historical wind speed and the forecasted wind speed. Many researchers turn to employ machine learning-based statistical methods, such as artificial neural networks (ANNs) [11],

* Corresponding author.
E-mail address: caddiexie@hotmail.com (Z. Xie).

https://doi.org/10.1016/j.enconman.2018.02.034
Received 27 November 2017; Received in revised form 1 February 2018; Accepted 8 February 2018

Abbreviations

Acronyms

AR	autoregressive model
ARMA	autoregressive and moving average model
ARIMA	autoregressive integrated moving average model
ANNs	artificial neural networks
BPNN	BP neural network
BMKLMOR	Bayesian multiple kernel learning for multi-output regression
C-MKL	a variant of CH-MKL when σ_i is same with each other
CH-MKL	heteroscedastic multi-kernel learning based correlation-aware multi-output regression
EEMD	ensemble empirical mode decomposition
Elman	Elman neural network
ELM	extreme learning machine
GP	Gaussian Process
GA	genetic algorithm
H-MKL	a variant of CH-MKL when $\boldsymbol{\Lambda} = \boldsymbol{I}_H$
LSSVM	least squares support vector machine
LSTM	long short-term memory neural network
MAE	mean absolute error
MAPE	mean absolute percentage error
MSVR	multi-output support vector regression
NWP	numerical weather prediction
PDF	probability density function
RMSE	root mean square error
SVM	support vector regression
VB	variational Bayesian

Symbols

$\mathscr{N}(\cdot)$	Gaussian distribution
$\mathscr{MN}(\cdot)$	matrix-variate Gaussian distribution
$\mathscr{IG}(\cdot)$	inverse Gamma distribution
$\mathscr{G}(\cdot)$	Gamma distribution
$\mathscr{IW}(\cdot)$	inverse-Wishart distribution
G,H	number of kernels and forecasting tasks
N,M	length of training set and dimension of $\boldsymbol{V}$
$\boldsymbol{\beta}$	weight vector of kernels
$\boldsymbol{\alpha}$	regression coefficient matrix
$\boldsymbol{Y}$,$\boldsymbol{X}$,$\boldsymbol{E}$	the target, input and error matrices
$\boldsymbol{M}$,$\boldsymbol{S}$,$\boldsymbol{\Omega}$	the mean, row and column covariance of $\boldsymbol{X}$
$\boldsymbol{K}_{\boldsymbol{\beta}}$	the sum of multiple kernel matrices
$\boldsymbol{I}_m$	$(m \times m)$ identity matrix
$\boldsymbol{\Xi}$,$\boldsymbol{\Lambda}$	the row covariance of $\boldsymbol{E}$, the column covariance of $\boldsymbol{Y}$
$\boldsymbol{\Omega}_\alpha$,$\boldsymbol{S}_\alpha$,$\boldsymbol{E}_\alpha$	the parameters of $\boldsymbol{\alpha}$'s posterior distribution
$\boldsymbol{\Omega}_Z$,$\boldsymbol{S}_Z$,$\boldsymbol{E}_Z$	the parameters of $\boldsymbol{Z}$'s posterior distribution
$\boldsymbol{\Omega}_V$,$\boldsymbol{S}_V$,$\boldsymbol{E}_V$	the parameters of $\boldsymbol{V}$'s posterior distribution
$\boldsymbol{\mu}_{\boldsymbol{\beta}}$,$\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$,	the parameters of $\boldsymbol{\beta}$'s posterior distribution
a_{new}^i,b_{new}^i	two parameters of σ_i's posterior distribution
c_{new}^g,d_{new}^g	two parameters of η_g's posterior distribution
e_{new},f_{new}	the parameters of γ's posterior distribution
g_{new},h_{new}	the parameters of τ's posterior distribution
$\boldsymbol{\Psi}_{new}$,ν_{new}	the scale matrix and the degree of freedom of $\boldsymbol{\Omega}$'s posterior distribution
$\boldsymbol{\Delta}$	the estimated task correlation by CH-MKL
a_0,b_0	the initialized parameters of a_{new}^i,b_{new}^i
c_0,d_0	the initialized parameters of c_{new}^g,d_{new}^g
e_0,f_0	the initialized parameters of e_{new},f_{new}
g_0,h_0	the initialized parameters of g_{new},h_{new}
$\boldsymbol{\Psi}_0$,ν_0	the initialized parameters of $\boldsymbol{\Psi}_{new}$,ν_{new}
$\boldsymbol{V}_0$,$\boldsymbol{Z}_0$	the initialized matrices of $\boldsymbol{V}$ and $\boldsymbol{Z}$

support vector machine (SVM) [12], Gaussian Process (GP) [13] and extreme learning machine (ELM) [14], to capture the nonlinear patterns hidden in wind speed data. Recently, some deep neural networks, including long short-term memory network (LSTM), begin to be applied in wind speed forecasting, and promising results are derived [4,15,16]. However, the structures of those ANNs play important roles in the final wind speed forecasting results [17], and ANNs often have the risk of local minima and over-fitting [18]. To overcome the above problems, some kernel-based nonlinear regression models such as SVM and GP are used [12,13]. However, their model parameters, including kernel parameter and regularization parameter, have great effects on the forecasting performance [19]. Though it takes less time to train an ELM than to train a SVM or GP, its forecasts are somewhat random due to the random assignment of its weights. Deep neural networks own strong fitting capabilities because the existence of a large number of parameters in the model. However, similar to ANNs, it is also hard to determine optimal structures for deep neural networks, more parameters need to be tuned, which will takes us much time.

The conventional statistical forecasting methods and the machine learning-based forecasting methods can model linear and nonlinear patterns hidden in wind speed times series, respectively. Naturally, hybrid models can be constructed by combining both of them to perform wind speed forecasting [20,21]. Results in [20,21] showed that the hybrid models outperformed single forecasting models. Another hybrid model integrates some algorithms, including signal processing and intelligent optimization algorithms, to enhance the forecasting capability of single forecasting models [22]. For instance, in [23], ensemble empirical mode decomposition (EEMD) was employed to preprocess the original wind speed data to reduce the adverse effects of noise and outliers on the final forecasting models, and genetic algorithm (GA) was used to tune the parameters in the wind speed forecasting model at the same time.

Recently, some new characteristics are discovered when conducting wind speed forecasting. Researchers turn to design new forecasting models based on these new characteristics. It was reported that the errors of wind speed forecasting didn't obey Gaussian distribution, but a Beta distribution [24,25]. Based on this finding, Beta noise based-SVM and Beta noise based-kernel ridge regression model were derived. The performances were tested on some real tasks [24,25]. Besides, heteroscedasticity is also observed in wind speed forecasting [26,27]. Considering this property, heteroscedastic support vector regression and heteroscedastic Gaussian process were developed in [26,27], respectively.

In this work, we focus on developing a multi-step ahead forecasting model based on the characteristics of wind speed forecasting. When conducting multi-step ahead forecasting, there are three commonly used strategies: iterated strategy, direct strategy and multi-output strategy [28]. For the iterated strategy, one-step ahead forecast is used as input to obtain the forecast in the following step [28]. Thus, it may suffer from the accumulated errors because the input variables include forecasts, rather than real observations [29]. As to the direct strategy, multi-step ahead forecasting is realized by adopting forecasting models to obtain each step ahead forecasts independently. It prevents those forecasting models with direct strategy from considering the dependency among different steps ahead forecasting tasks [28,29]. To overcome the above shortcomings, multi-output forecasting models were proposed, including neural networks [30], multi-output support vector regression (MSVR) [31] and Bayesian multiple kernel learning