Solar Energy 159 (2018) 97–112

Contents lists available at ScienceDirect

Solar Energy

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

Interval prediction of solar power using an Improved Bootstrap method

Kaiwen Li[a], Rui Wang[a,*], Hongtao Lei[a], Tao Zhang[a,b], Yajie Liu[a], Xiaokun Zheng[a]

[a] College of Information Systems and Management, National University of Defense Technology, Changsha 410073, PR China
[b] State Key Laboratory of High Performance Computing, National University of Defense Technology, Changsha 410073, PR China

ARTICLE INFO

Keywords:
Solar power
Prediction intervals
Bootstrap
Envelope

ABSTRACT

The integration of solar energies into power grid requires accurate prediction of solar power. While most previous literature is focused on how to improve the accuracy of point forecast, we consider constructing the prediction interval (PI) for solar power which is more appropriate for its nature of high variability. Traditional theoretical approaches of constructing PIs always require the assumption that forecast errors are normally distributed with zero mean. However, this assumption can be easily invalid for solar power data. In this work, an Improved Bootstrap method is proposed to improve the traditional theoretical approaches. It is especially designed to provide PIs for solar power and the problem of invalid assumption about forecast errors can be addressed. The proposed methodparison with three different types of novel PI methods. With interval width and coverage probability as evaluation measures, our method achieves a more than three times lower interval width than other methods while the coverage probability can be still guaranteed. Two-year photovoltaic data of the University of Queensland is used to validate the methodology and different prediction time frames of 5 min, 30 min, 1 h, 2 h and 6 h are applied.

1. Introduction

As most of the fossil resources are on the verge of depletion, the demand for renewable energy sources has shown a steady increase in recent years. Meanwhile, climate change problems, i.e. global warming and the rise of sea level, call for concentrated efforts to reduce the emission of CO_2. Compared with fossil resources, renewable energy sources are abundant, renewable and environmental friendly. In recent years, solar energies are one of the most promising renewable sources and have high penetration in energy market (Raza et al., 2016).

However, solar energies also face a lot of challenges due to its uncertain nature. The high variability and uncertainty may lead to various problems for the reliable and economic operation of power grid. In order to ensure the stability of power grid, accurate forecast of the solar power is needed. Most of the previous literature is focused on point forecast, which is to predict a single value in the future. However, the interval prediction, which can obtain the upper and lower bound of a future value, is more suitable for solar power due to its high variability. It is more applicable for systems requiring risk management like electricity production (Torgo and Ohashi, 2011). Predicting an interval offers additional variability information than just predicting a single value. When knowing the range of target point, better energy management can be made to reduce the operation cost and improve stability of power grid system.

There are mainly two types of approaches to estimate prediction intervals in the literature. First type is the theoretical approach. For this approach, theoretical interval is calculated based on the assumption that forecast errors follow a determined distribution with zero mean, usually the normal distribution (Yun and Scholtes, 2014). However, in real world where data always involves complex processes, it is hard to ensure the assumption can be fulfilled. The theoretical prediction interval may behave poorly if the aforementioned assumption is not valid. As alternatives, another type of approaches have been proposed with no need of consideration of the forecast error distribution. Empirical approach is a typical one of such approaches. And this type of approaches is claimed to achieve robust performance for the construction of PIs (Yun and Scholtes, 2014).

In the field of solar energy, there are some studies on constructing prediction interval for solar irradiance. However, the interval prediction for solar power is little studied. The aim of our work is to develop a suitable PI construction method especially designed for the solar power. As the solar power data is complex, asymmetric and there might be multiple data patterns within a large solar dataset, such as data of clear day and cloudy day, the forecast error of the derived forecast model may not obey a normal distribution with zero mean. The theoretical approach might show a poor performance for the construction of solar power PIs. In this work, an Improved Bootstrap method is proposed to improve the traditional theoretical approaches. This method can

* Corresponding author.
E-mail addresses: kaiwenli_nudt@foxmail.com (K. Li), ruiwangnudt@gmail.com (R. Wang), hongtaolei@aliyun.com (H. Lei), zhangtao@nudt.edu.cn (T. Zhang).

http://dx.doi.org/10.1016/j.solener.2017.10.051
Received 19 August 2017; Received in revised form 13 October 2017; Accepted 14 October 2017

address the aforementioned problems and provide better prediction intervals.

In the remainder of this paper, the basic theory of the construction of PI and related work in recent years are introduced in Section 2. Solar data used in this work and the selection of model inputs are introduced in Section 3. Section 4 highlights the problems of applying traditional Bootstrap method to construct PIs for solar power data. The Improved Bootstrap method is developed accordingly. Section 5 validates the efficiency of the proposed method by comparison with different types of PI methods. The paper is concluded by discussion and future work in the last section.

2. Literature review

The problem of constructing prediction intervals has been studied mainly by two means: theoretical approaches and non-theoretical approaches (Yun and Scholtes, 2014). In order to understand the whole work, the basic theory of constructing theoretical PIs is first illustrated in this section. Moreover, recent literature on theoretical and non-theoretical approaches are introduced.

2.1. Basic theory

Theoretical approaches for the construction of PIs are based on the assumption that forecast models are unbiased and forecast errors follow a determined distribution with zero mean. The basic theory of the theoretical approach is introduced as follows.

Given a finite number of data points

$$\{(x_i,t_i)\}_{i=1}^{n}, \tag{1}$$

a regression task is to estimate an underlying mathematical function between input variables x and output variables t. It can be modeled by

$$t_i = y_i + \epsilon_i \tag{2}$$

where t_i is the ith observed target (totally n targets). y_i is the true regression mean. As the regression results cannot fit all the observed values, there are always errors between t_i and y_i. The error ϵ_i with zero mean is usually called noise. In practice, a specific regression model $\hat{y}_i$ is built to estimate the true regression mean y_i. However, in most instances, $\hat{y}_i$ is not equal to the true regression because of the model misspecification and parameter estimation errors. According to this, we have

$$t_i - \hat{y}_i = [y_i - \hat{y}_i] + \epsilon_i \tag{3}$$

PIs deal with the difference between the observed values t_i and the predicted values $\hat{y}_i$, as shown in the left-hand side of (3). Confidence intervals (CIs) quantify the uncertainty between the prediction $\hat{y}_i$ and the true regression y_i, as shown in the first term in the right-hand side of (3). We see from (3) that a prediction interval necessarily encloses the corresponding confidence interval.

Assuming the two noise components $[y_i - \hat{y}_i]$ and ϵ_i in (3) are independent, the variance associated with the difference between observed values and predicted values will be:

$$\sigma_i^2 = \sigma_{\hat{y}_i}^2 + \sigma_{\epsilon_i}^2 \tag{4}$$

$\sigma_{\hat{y}_i}^2$ is the measure of model variance, mainly due to the different model parameters or using different training data. $\sigma_{\epsilon_i}^2$ is the measure of noise variance.

Assuming a normal distribution with zero mean of the forecast errors, then the true value is supposed to fluctuate around the forecast value. In detail, the $(1-\alpha)100\%$ prediction interval can be constructed as

$$\left(\hat{y}_i \pm z^{1-\frac{\alpha}{2}}\sigma_i\right) \tag{5}$$

where $(1-\alpha)100\%$ is the confidence level and $\hat{y}_i$ is the point forecast. The $(1-\alpha)100\%$ confidence level means that the prediction values are supposed to lie within the interval with a prescribed probability $(1-\alpha)100\%$.

2.2. Recent work

In the literature, several methods have been developed to construct he theoretical PIs based on the aforementioned assumption about forecast errors.

Delta method (Hwang and Ding, 1997; VlEAUX et al., 1998) has been proposed to construct PIs using the neural network (NN) as forecast model. PIs can be constructed by applying asymptotic theories to the linearized NN model (Seber and Wild, 2003). The assumption that errors are normally distributed is required. It suffers a high computational cost. Mean-variance estimation (MVE) method (Nix and Weigend, 1994) is another method proposed by Nix and Weigend for construction of PIs. NN model is used to estimate the distribution characteristics of the forecast target. After obtaining distribution of the target, PIs can be constructed accordingly. Unlike delta method who uses a fixed variance, MVE assumes a non-constant gaussian variance. The computational cost for this method is negligible but it suffers low coverage probability (Ding and He, 2003). The Bayesian technique (Kothari and Oh, 2001; Mackay, 1992) is also developed for construction of NN-based PIs. A regularized cost function is applied to train the NN model. And a better generalization power is acquired than other networks. Similar to the delta technique, it is computationally demanding as it requires calculation of the Hessian matrix for the construction of PIs.

Bootstrap method (Beran, 1992; Heskes, 1997) is one of the most frequently used technique in the literature for construction of confidence intervals (CIs) and PIs. For the process of constructing PIs, there are mainly two types of approaches: using theoretical way (as illustrated above) and using percentiles way. The percentile bootstrap does not require the statistic assumption of forecast errors. However, the width of intervals provided by the percentile bootstrap are often to be too narrow (Scheiner and Gurevitch, 2014). Moreover, bootstrap method can also be divided into parametric and non-parametric bootstrap respectively (Efron and Tibshirani, 1993). The parametric bootstrap requires that original data is also supposed to be normally distributed, not only the forecast errors. And the non-parametric bootstrap has no assumption of the data distribution, just requiring the normal distribution of forecast errors. The detailed description of bootstrap will be illustrated in Section 4. The advantage of bootstrap method is its simplicity and ease of implementation. In addition, it has been researched to outperform delta method and Nix-Weigend method in the literature (Heskes, 1997). Due to the outstanding performance of the Bootstrap method among the theoretical approaches, it is selected to be improved in this work for a better prediction of the solar power.

Without limit of the distribution assumption about forecast errors, several types of methods have been proposed. First, Empirical approaches are popular in the literature, which involve two types. The first type is using empirical residual errors to construct PIs. Forecast errors at different lead times can be obtained by applying the forecast model to the same data used for fitting the forecast model. On this error data, some non-parametric methods are applied to construct PIs, such as Chebyshev's inequality (Gardner, 1988), kernel density estimators (Wu, 2012), and semi-parametric techniques such as quantile regression (Taylor and Bunn, 1999). As the residual errors are usually lower than out-of-sample errors, the PIs constructed may be too narrow. Thus, another type of empirical approach is based on the out-of-sample errors. PIs are constructed on the errors generated with the data never used before, called out-of-sample. And this approach was initially developed by Williams and Goodman (1971), which has been applied increasingly then.

There are some other novel non-theoretical approaches proposed recently. A method called Lower Upper Bound Estimation (LUBE) (Khosravi et al., 2011b) is proposed with no need of the assumption