Energy 154 (2018) 143–156

Contents lists available at ScienceDirect

Energy

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

Short-term power load probability density forecasting based on Yeo-Johnson transformation quantile regression and Gaussian kernel function

Yaoyao He [a, b, *], Yaya Zheng [a, b]

[a] School of Management, Hefei University of Technology, Hefei 230009, China
[b] Key Laboratory of Process Optimization and Intelligent Decision-Making (Hefei University of Technology), Ministry of Education, Hefei 230009, China

ARTICLE INFO

Article history:
Received 28 June 2017
Received in revised form
21 March 2018
Accepted 13 April 2018
Available online 18 April 2018

Keywords:
Load forecasting
Yeo-Johnson transformation quantile regression
Probability density forecasting
Gaussian kernel function
Quantile function crossing

ABSTRACT

Penetration of renewable resources into power systems, such as wind and solar power, has significantly grown the complexity and level of uncertainty in both power generation and demand sides, which are highly desirable to exploit more advanced methods to address the uncertainties. The probability density forecasting method using quantile regression can describe probability distributions of future power load. However, existing quantile regression probability density forecasting methods may encounter embarrassing cross phenomenon, affecting the effectiveness of probability density forecasting. To avoid the crossing issue, this study proposes a probability density forecasting method based on Yeo-Johnson transformation quantile regression using Gaussian kernel function. Gaussian kernel density estimation using a rule of thumb bandwidth is innovatively hybridized with Yeo-Johnson transformation quantile regression for short-term power load probability density forecasting. The evaluation metrics for forecasting errors and prediction interval are adopted to carry out a comprehensive study on load uncertainty handling. One-hourly historical load data of August 2014 in summer and December 2014 in winter from Ottawa, Canada are used to evaluate the performance of proposed model. The results show that the proposed method not only efficiently avoids the quantile crossing problem but also obtains smooth probability density curves and more accurate forecasting results.

1. Introduction

Load forecasting has a definitive impact on the operation of the power system. Various operation decisions depend on load forecasting, such as price and income elasticity, energy transfer scheduling, unit commitment and load dispatch for generators [1]. The deregulation and free competition of power systems have increased in many countries, making the load forecasting more important than before. The electricity supply needs to be determined by effective forecasting models as the electricity is hard to be stored. Forecasting errors can deliver the information about either unnecessary shortages or overproduction of electricity for future load planning. System operators depend on reliable forecasting results to deploy the demand response (DR) programs, keeping in mind simultaneously that uncertainty is a key issue to most decisions [2,3]. In general, load forecasting can be categorized into short-term, medium-term, and long-term in accordance to the forecast horizon [4]. The short-term load forecasting (STLF) is a core issue for daily operations, security analysis of system and maintenance scheduling. However, it is increasingly difficult to forecast short-term power load, due to the variability and non-stationarity of load series. Meanwhile, in smart grid, penetration of renewable sources into power systems on the one hand potentially enhances the overall system performance in terms of economic and security [5]. On the other hand, it increases the uncertainty in both power generation and demand sides, which forms huge challenge to power load forecasting [6,7]. Thus, advanced models need to be used for quantifying the uncertainty of STLF.

The uncertainty of load forecasting is expressed as the probability distribution around a point forecasting value. According to forecasting outputs, load forecasting can be divided into point forecasting, interval prediction [8] and probability density

* Corresponding author. School of Management, Hefei University of Technology, Hefei 230009, China.
E-mail addresses: hy-342501y@163.com (Y. He), 1020436426@qq.com (Y. Zheng).

https://doi.org/10.1016/j.energy.2018.04.072
0360-5442/

forecasting [9]. Most of studies performed on STLF focus on point forecasting. However, as for as point forecasting methods, one target value can be provided by one predicted point, and it is unable to measure the uncertainties in power systems [10,11]. Prediction intervals (PIs) are useful tools to quantify the uncertainties related with forecasts. A PI is the forecasting of an interval in which the future actual value will fall, with a certain probability named as the confidence level [12]. As the most complete prediction method, probability density forecasting not only effectively addresses the uncertainties associated with datasets, but also has the ability to construct complete conditional probability density curves of future loads. However, many assumptions and mathematical derivations are usually defined in advance and hinder the application of probability density forecasting [13]. Hence, probability density forecasting is not widely studied.

In recent years, many advanced models are performed on STLF. Artificial neural network (ANN) [14,15], and support vector regression (SVR) [16] are two most popular ones. ANN is capable of solving the probabilistic prediction problems using the mean and variance of a Gaussian distribution when sufficient observed data are supplied [14]. Reference [17] forecasted the daily load by ANN and the output of ANN forecaster was used for determining the need to load reduction. In Ref. [10], three different techniques (i.e., errors output, resampling and multilinear regression) were investigated into STLF for establishing confidence intervals by means of ANN models. Guan et al. [18] applied wavelet neural network (WNN) to STLF and PI estimation. The WNN was trained by hybrid Kalman filters for very STLF. However, ANN suffers from many disadvantages in many real-world case studies. For example, it depends on a large number of controlling parameters and has difficulty in obtaining solution. Considering the complexity and potential nonlinearity of power load, SVR [1,19] has been proposed to handle these problems, which has led it to become one of effective methods due to the perfect performance in electricity consumption forecasts [20]. Wang et al. [21] established a hybrid load forecasting model combining differential evolution (DE) and SVR to choose the appropriate parameters. As a kernel-based method, SVR has the advantage of mapping the input data from a low dimensional space to a high-dimensional feature. It can flexibly convert nonlinear regression into linear regression without assuming particular functional forms [22]. However, the forecasting performance of SVR mainly depends on the degree of its three parameter values [23]. So far, there are still no fine solutions to be used to determine the optimal parameters of SVR. Furthermore, the ANN and SVR models can not completely quantify the uncertainties in power systems, and only limited point forecasts and interval predictions are acquired. Hence, it is necessary to explore the appropriate methods to quantify the probability distribution of predictive values.

Different from the above methods, quantile regression (QR) can directly estimate the point values at different quantiles. QR has the advantage of providing predictive values throughout the range of quantiles without assuming the parametric form of distribution functions [24,25]. QR is capable of solving a multitude of complicated problems. Thus, it has been widely used in power load forecasting in the past decade [26–28]. However, one disturbing problem with quantile regression is that quantile curves obtained can cross each other, leading to an invalid distribution for the response. For example, the forecasted 90th percentile of the response is smaller than the 85th percentile. Theoretically, the estimated quantile function crossing violates the basic principle that distribution functions should be monotone increasing. This disturbing cross phenomenon reduces the estimation accuracy and makes it difficult to interpret the resultant models [29]. This phenomenon has been observed in some multiple conditional quantile regressions. In order to overcome the quantile function crossing problem, Chernozhukov et al. [30] estimated non-crossing quantile via a monotonic rearrangement of original non-monotone function. In Ref. [31], the author employed constraints on kernel coefficients to guarantee the estimated conditional quantile functions never cross each other.

Considering the advantages of QR, many researchers have incorporated artificial intelligence models into QR to construct a probability density forecasting method, such as support vector quantile regression (SVQR) [12,32,33] and quantile regression neural network using triangle kernel function (QRNNT) [13,34]. In Ref. [35], the author combined the QRNN with kernel density estimation to construct probability density forecasting model of wind power. These models can quantify the prediction uncertainty and solve complicated nonlinear problems. The results of probability density forecasting based on QRNNT and SVQR have shown perfect performance in practical applications. However, these methods can not avoid the quantile function crossing problem. A successful approach proposed by Cole and Green [36,37] assumed that a suitable transformation would acquire normality of the response along with the location and scale, fully determining the distribution of quantile functions. They described in detail the method of maximum penalized likelihood to estimate three parameters called $\lambda, \mu, and\ \sigma$. This technique was named LMS method from which the method derives the name of three parameters (i.e., the first letters of $\lambda, \mu, and\ \sigma$ starting with 'L-M-S' respectively, hence its name). LMS method is one of the most popular quantile regression methods due to its simplicity and flexibility. The method naturally falls within a penalized likelihood framework, and consequently allows for considerable flexible because all three parameters may be modeled by cubic smoothing splines. Considering the superiority of the LMS method, this study proposes a power load forecasting method based on Yeo-Johnson transformation quantile regression (YJQR) which is a new version of LMS method.

As the most complete forecast method, probability density forecasting can effectively represent uncertainties as the probability distribution around a point forecast. However, a single YJQR is hard to describe the regularity and probability of future power load satisfactorily. To overcome this problem, we take advantage of YJQR technology and kernel density estimation (KDE) theory to construct a probability density forecasting method based on YJQR and Gaussian kernel density estimation (YJQRG). Due to the strong generalization ability of Gaussian kernel function, we adapt it to construct probability density functions that have significant effects on the distribution of response variables. Rule of thumb is a reliable bandwidth selection method and further enhances the smoothness and continuity of probability density curves through the integration of obtained quantile functions. The proposed hybrid method not only implements the conditional probability density forecasting, but also improves the estimation efficiency by applying Gaussian kernel function and rule of thumb to estimate the scope of prediction intervals. One-hourly practical data collected from Ottawa (OTT), Canada were employed to power load forecasting based on YJQRG.

In recent years, penetrations of a large number of distributed generations into OTT power systems increase the uncertainties in both generation and demand sides. Among these distributed generations, wind and solar power generations have important positions. As renewable energy sources, wind and solar power generations are intermittent and volatile by nature [38]. The uncertainties increase the difficulties of power system operations and are urgent to be addressed. This paper constructs YJQRG to quantify the uncertainties in OTT power systems and improves the accuracy of STLF. Considering the seasonal differences of OTT, we choose two months of data in August 2014 and December 2014. From the