Applied Energy 205 (2017) 790–801

Contents lists available at ScienceDirect

Applied Energy

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

Short-term electricity demand forecasting using autoregressive based time varying model incorporating representative data adjustment

D.H. Vu*, K.M. Muttaqi, A.P. Agalgaonkar, A. Bouzerdoum

School of Electrical, Computer and Telecommunications Engineering, University of Wollongong, Wollongong, New South Wales, Australia

HIGHLIGHTS

- An autoregressive based time varying model is proposed to forecast electricity demand in short-term period.
- Sine and cosine functions are selected as time varying function of the model.
- A demand based adjustment procedure is proposed to mitigate the influences of daylight saving and holidays.
- A case study is conducted with the aid of a dataset acquired from NSW, Australia to validate the proposed model.
- A comparative analysis has been conducted to illustrate the effectiveness of the demand based adjustment procedure.

ARTICLE INFO

Keywords:
Electricity demand forecasting
Autoregressive based time varying model
Similar-day-replacement technique

ABSTRACT

This paper presents the development of an autoregressive based time varying (ARTV) model to forecast electricity demand in a short-term period. The ARTV model is developed based on an autoregressive model by allowing its coefficients to be updated at pre-set time intervals. The updated coefficients help to enhance the relationships between electricity demand and its own historical values, and accordingly improve the performance of the model. In addition, a representative data adjustment procedure including a similar-day-replacement technique and a data-shifting algorithm is proposed in this paper to cultivate the historical demand data. These techniques help cleanse the raw data by mitigating the abnormal data points when daylight saving and holiday occur. Consequently, the robustness of the model is significantly enhanced, and accordingly the overall forecasting accuracy of the model is considerably improved. A case study has been reported in this paper by acquiring the relevant data for the state of New South Wales, Australia. The results show that the proposed model outperforms conventional seasonal autoregressive and neural network models in short term electricity demand forecasting.

1. Introduction

Load forecasting is crucial in electricity network operation as it can provide critical information for not only maintaining the balance between the load and generation, but also planning the expansion of power grids. While over-forecast may lead to starting-up unnecessary power generator commitments, under-forecast could lead to purchases of expensive peaking demand due to lack of generation preparation. Consequently, more accurate forecast of electricity demand can help to harvest the generation economically and gain more benefit in operating the system [1]. With the trend of energy internet development, the demand forecasting is an essential part in building the business model to maximise the profit and efficiency of the network [2]. It is necessary to obtain and analyse the user patterns and prepare requisite generation for the peak demand. Furthermore, reliability and quality of the power supply are enhanced significantly with the credibility of demand forecasting. For example, with an accurate demand prediction, the generation can be scheduled to meet demand in tough occasions, so unexpected and precarious load shedding events can be limited and avoided effectively.

Recently, a considerable literature has been dedicated around the theme of load forecasting [3], and numerous models and techniques have been employed for forecasting the load in short timeframes [4]. These techniques can be categorised as either classical statistical models or machine learning models. While the former has advantages in transparent interpretation, the machine learning models have strength in the ability to model nonlinearities. A typical representative of machine learning is neural network, which was used in [5] to forecast the

* Corresponding author.
E-mail addresses: dhv972@uowmail.edu.au (D.H. Vu), kashem@uow.edu.au (K.M. Muttaqi), ashish@uow.edu.au (A.P. Agalgaonkar), a.bouzerdoum@uow.edu.au (A. Bouzerdoum).

http://dx.doi.org/10.1016/j.apenergy.2017.08.135
Received 15 March 2017; Received in revised form 7 July 2017; Accepted 12 August 2017

load thanks to its ability to model the nonlinear components. In this paper, the neural network model is trained using the deviation of load and similar load profile days, which are selected using the weighted Euclidean norm. To boost the performance of neural network model, wavelet technique was employed in [6,7] for components decomposition of the input variables. The decomposed components are orthogonal, and they help provide more supportive information to neural network and accordingly increase the performance of the forecasting model. In addition, the authors in [8] used the neural network in combination with fuzzy function, while the authors in [9] employed particle swarm optimisation algorithm to select the neural network parameters. Recently, multiple neural networks have been combined in [10] and [11] to tackle the demand forecast as the combined model helps reduce uncertainty in the forecasting results. However, the neural networks can be considered as a black box in the training process, so the interpretability of the model parameters may be limited. Also, if the parameters of the networks are not properly selected, the obtained model may not be stable [12].

Unlike neural network models, the statistical models are transparent so they are more interpretable. The representatives of the statistical models include regression model and Box-Jenkin models. The regression model was used in [13] to forecast the electricity demand due to its simplicity. In order to improve the forecasting performance, the regression model in [14] was used to model the trend only, and this result was incorporated with the seasonal components from seasonal exponential method to obtain the forecasting results. The functional regression model was developed in [15] using the epi-splines function, which helps capture the non-weather dependent load pattern in the forecasting model, while principle component analysis technique was used in [16] to help extract more information from the input variables of the model. Other techniques such as Kalman filter [17,18] and support vector machine [19,20] were demonstrated to help improve the forecasting results. Among statistical methods, autoregressive (AR) model can be used to forecast the demand in short term period effectively [21]. This model is one type of Box-Jenkins models [22], and it can be extended to autoregressive moving average (ARMA) model, which helps to improve the ability in capturing the characteristics of the demand [23]. With regard to ARMA models, the integration part can be included to create a model that can deal with the datasets which are non-stationary [24,25]. In addition, the seasonality can be included into the ARMA based model to capture the seasonal variations of the electricity demand in a seasonal autoregressive integrated moving average (SARIMA) model [26]. These models are all autoregressive-based static models, whose coefficients are fixed at pre-set values. The fixed-coefficients may constrain the flexibility of the model since adaptability with the evolution of the system is missing. As a result, the model's performance may be not as good as expected when comparing with other forecasting models [27,28]. In order to improve the performance of the AR based model, multiple ARMA models have been proposed in [29]. In this paper, the authors have built different ARMA models to forecast load on weekdays and weekends. The coefficients of each ARMA model are then updated with the error learning weight in each forecasting step. This technique helps improve the forecasting ability of the ARMA model but this development may increase the complexity of the model since many sub-models need to be built to track the demand for different hours [30] and different seasons [31].

In the proposed research study, an autoregressive based time varying (ARTV) model is developed to forecast electricity demand in short term period. In this model, the associated coefficients are updated at a pre-set time interval following a pre-defined time varying function. This function determines hourly and seasonal information which are included into the forecasting model, and these inclusions result in the flexibility of the coefficients in different seasons, and finally leads to further improvement in model performance without reliance on any sub models. Also, in this paper, a demand based adjustment procedure is proposed to mitigate the inherent abnormalities in electricity demand data related to the influences of daylight saving and holidays. This mitigation process contributes significantly to the accuracy improvement and robustness enhancement of the forecasting results. With the aid of the electricity demand data from the State of New South Wales (NSW), Australia, the model has been validated, and the results have shown that the time varying model significantly improves the forecasting performance compared to conventional AR model and it outperforms three bench mark models including naïve model, autoregressive model and neural network model.

The contribution of the paper can be summarised as follows.

- An ARTV model has been developed to forecast the short-term load forecasting by incorporating a time varying component to the AR based model. As a result, the obtained ARTV model has the ability to adapt its coefficients with time evolution and subsequently results in significant performance improvement without the reliance on many sub models.
- An investigation into the impacts of special events including daylight saving and holiday on training dataset has been conducted, and the results show that these events can create misleading information in training dataset and subsequently cause significant deterioration in forecasting performance. Consequently, a demand based adjustment procedure has been proposed to cleanse the training dataset and mitigate the misleading information from those incidents. After this procedure, the robustness of the model is greatly enhanced and the accuracy of the model is significantly improved.
- A case study has been carried out with the aid of the data from the state of New South Wales (NSW), Australia, and the results show that the proposed ARTV model is highly robust and accurate in short-term demand forecasting with the overall forecasting performance (MAPE) in entire year of 2015 being at 0.621%. Furthermore, a comparative analysis has been conducted and the results indicate that the proposed ARTV model strongly outperforms three benchmark models including naïve model, AR model, and neural network model.

The rest of the paper is organised as follows: Section 2 presents the development of the ARTV model. Section 3 introduces the demand based adjustment procedure to mitigate the abnormal data points in the historical data. Section 4 highlights some experimental results and model validation. Section 5 provides the concluding remarks of the proposed research study.

2. Autoregressive based time varying model

2.1. Autoregressive models

Autoregressive (AR) model is one type of the Box-Jenkins models [22]. This model is widely used in statistical applications especially in electricity demand forecasting [21]. The idea of autoregressive model is to link the forecasting value to the historical data via a regressive equation. The mathematical representation of AR model for variable $y(k)$ is given as in (1).

$$y(k) = \sum_{i=1}^{m} a_i \times y(k-l_i) + e(k) \tag{1}$$

where m is the total number of lags included in the equation, a_i is the coefficient, $y(k-l_i)$ is the time lag l_i of variable $y(k)$ and $e(k)$ is the residual.

It is noted that the autoregressive model which is presented in (1) uses fixed coefficients. These coefficients are generalised over a training dataset and they remain constant for an entire forecasting period. This generalisation causes mismatch in linking between demand and historical data for different hours, and subsequently results in poor accuracy in forecasting results. Consequently, time varying coefficients