Advanced Engineering Informatics 30 (2016) 211–217

Contents lists available at ScienceDirect

Advanced Engineering Informatics

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

Full length article

Extended modeling procedure based on the projected sample for forecasting short-term electricity consumption

Che-Jung Chang [a,*], Jan-Yan Lin [b], Meng-Jen Chang [b]

[a] Department of Management Science and Engineering, Business School, Ningbo University, No. 818, Fenghua Road, Ningbo City, Zhejiang Province 315211, China
[b] Department of Business Administration, Chung Yuan Christian University, No. 200, Chung-Pei Road, Chung-Li District, Taoyuan City, Taiwan 32023, ROC

ARTICLE INFO

Article history:
Received 14 August 2015
Received in revised form 22 January 2016
Accepted 22 March 2016
Available online 31 March 2016

Keywords:
Forecasting
Small data set
Latent information
Electricity consumption

ABSTRACT

Effectively forecasting the overall electricity consumption is vital for policy makers in rapidly developing countries. It can provide guidelines for planning electricity systems. However, common forecasting techniques based on large historical data sets are not applicable to these countries because their economic growth is high and unsteady; therefore, an accurate forecasting technique using limited samples is crucial. To solve this problem, this study proposes a novel modeling procedure. First, the latent information function is adopted to analyze data features and acquire hidden information from collected observations. Next, the projected sample generation is developed to extend the original data set for improving the forecasting performance of back propagation neural networks. The effectiveness of the proposed approach is estimated using three cases. The experimental results show that the proposed modeling procedure can provide valuable information for constructing a robust model, which yields precise predictions with the limited time series data. The proposed modeling procedure is useful for small time series forecasting.

1. Introduction

Energy is a vital strategic resource that affects the national economy and social development [1]. Economic growth and industrialization rapidly increase energy consumption and production. Therefore, a nation's energy policy is crucial, because it not only guides the development of a country but also affects the operating environment of various industries [2]. Electricity is a form of energy that is arduous to store [3], and considerable evidence supports a causal relationship between economic growth and electricity consumption [4].

Because of the large amount of capital investment and lengthy construction time required in electricity systems expansion planning, an incorrect direction of development causes a dramatic effect. To advance the economic growth and fulfill future power requirements, forecasting electricity consumption effectively is essential. However, effective forecasting has become a challenge to overcome. Therefore, forecasting the future electricity consumption correctly and scientifically to manage power systems is crucial [5].

Several methods have been used to forecast electricity consumption over the last few decades [6–10]. These forecasting methods can be approximately classified into three categories: causal models, time series analysis, and artificial intelligence approaches. Causal models probe the relationships among multiple variables and assume that the variations in dependent variables can be explained by independent variables; specifically, historical data are used to establish a multivariate model for dependent variable forecasting [11]. The forecasting accuracy of a causal model depends on the selection of independent variables. If the selected independent variables cannot effectively explain the variation in the dependent variables, an inaccurate forecast is produced.

Time series models include linear regression and autoregressive integrated moving average analysis. Time series models require only historical observations to construct a model for forecasting the development of data trends [12] and are commonly used in forecasting energy demand. However, they require high quantities of samples for accurate forecasting.

Artificial intelligence approaches include data mining techniques, artificial neural networks, and heuristic algorithms. They are often used to solve forecasting problems and attain extremely high forecasting performance [13]. However, the forecasting results depend on the number of training samples and their representativeness; these modeling restrictions must to be overcome.

* Corresponding author.
E-mail addresses: r3795102@nckualumni.org.tw (C.-J. Chang), rock@cycu.edu.tw (J.-Y. Lin).

http://dx.doi.org/10.1016/j.aei.2016.03.003
1474-0346/

In all the aforementioned methods, the sample size is a key element that affects the forecasting performance and limits the applicability of the forecasting approach to certain situations; forecasting the energy demand in rapidly developing countries is a positive example of it. Although it can collect a large amount of historical observations, they usually deviate considerably from the real increasing trend in electricity consumption. Because electricity consumption typically exhibits an exponential trend, common forecasting methods using large quantities of historical data, such as basic time series approaches, are unsuitable [14]. Therefore, it is beneficial to develop a new modeling procedure by using small data sets for forecasting the electricity consumption.

The difficulty of small-data-set learning tasks is due to the limited samples being unable to completely reflect all characteristics of a population [15]. To overcome this particular forecasting problem, some studies have adopted virtual sample generation (VSG) techniques to enhance the learning stability for constructing robust and exact models [16–20]. These VSG approaches have been used in many fields, such as manufacturing, engineering, and medicine [21–24]. However, the process of generating artificial samples usually does not consider the relationship among dependent attributes, limiting its usefulness. Time series data constitute one typical example of dependent attribute data. In such data, correlations exist between the developing trends of observations and the order of observations. Hence, VSG is not applicable to time series data; the generated virtual samples cannot effectively improve the modeling performance because the relationship between a datum and time is not considered.

For solving this learning problem, this study proposes a procedure that involves first employing the latent information (LI) function proposed by Chang et al. [25] for analyzing data to extract the concealed information. Next, we develop the projected sample generation for combining the LI values and original data to extend the data set and thus enhance the forecasting accuracy of a back propagation neural network (BPNN).

The main purpose of the projected samples is to provide additional information for enhancing the learning stability. VSG techniques improve the learning performance by extending the sample size. Their similarity between the projected samples and VSG is that they increase the information input in the learning process; however, the difference is the type of information input. The sample pattern generated using VSG is the same as that of the original sample, whereas that of the projected sample is different. To increase the information input, the projected sample generation changes the dimension of the independent variable.

To verify the effectiveness of the proposed approach, this study first used electricity consumption data collected from the Asia–Pacific Economic Cooperation (APEC) energy database. Moreover, two additional cases, data on the wafer-level packaging (WLP) process and monthly demand for thin film transistor liquid crystal display (TFT-LCD) panels, were examined to further verify the performance of the proposed approach. The experimental results showed that the proposed modeling procedure based on the LI function is an appropriate technique for small-data-set forecasts because of its ability to provide valuable information and precise forecasts with limited time series data.

The remainder of this paper is organized as follows. The LI function, projected sample generation, and modeling procedure are introduced in Section 2. Section 3 provides comparisons among forecasting methods. Finally, the conclusion is presented in Section 4.

2. Methodology

The learning procedure of a forecasting model is ineffective when the sample size is small because it provides insufficient information. Therefore, this study developed a novel modeling procedure, called projected sample generation, to enhance the short-term forecasting performance with limited time series data. The concept and implementation of the proposed approach are described in this section.

2.1. LI function

Previous studies have reported that increasing the information content enables obtaining a sufficient sample for effective learning and thus attaining a more stable forecasting result. Therefore, this study employed the LI function for analyzing data behavior and extracting hidden information to facilitate discovering new knowledge by using small data sets.

The LI function was proposed by Chang et al. [25], and its main concept is to appropriately expand the margins of data by using four indices to fill the data gaps, where the extent of the range is determined by the sample size. Specifically, for a high number of observations, the data profile becomes clearer because of the high amount of information, and substantially extending the data range is not necessary. Conversely, if the information is insufficient, the extent of the range must be increased. The degree of extension can be determined using the range divided by the number of samples, and the ratio of leftward or rightward extension is determined according to the Skewness. The extended boundaries are called the upper bound (UB) and lower bound (LB), which are then combined with the central tendency (CT) to jointly construct the LI function. The range of the LI values lies between 0 and 1, representing the likelihood of occurrence of the potential data. The complete procedure for formulating the LI function is described as follows:

1. For an n-periods time series data set $X = \{x_1, x_2, \ldots, x_n\}$, let $x_{\min}$ and $x_{\max}$ be the element in X with the minimal and maximal values, respectively. Then calculate the range R by using Eq. (1).

$$R = x_{\min} - x_{\max} \tag{1}$$

2. Determine the CT using Eq. (2).

$$\mathrm{CT} = \frac{\sum_{i=1}^{n} i x_i}{\sum_{i=1}^{n} i}, \quad i = 1, 2, \ldots, n \tag{2}$$

3. Determine the central location (CL) of the existing data by using Eq. (3).

$$\mathrm{CL} = \frac{x_{\min} + x_{\max}}{2} \tag{3}$$

4. Determine the number of elements in the subset comprising data with values greater than the CL and denote it by $|X^+|$; determine the number of elements in the subset comprising data with values less than the CL and denote it by $|X^-|$.
5. Compute the increasing tendency (IT) and decreasing tendency (DT).

$$\begin{cases} \mathrm{IT} = \frac{|X^+|}{|X^+|+|X^-|} \\ \mathrm{DT} = \frac{|X^-|}{|X^+|+|X^-|} \end{cases} \tag{4}$$

6. Employ the IT and DT to asymmetrically expand the domain range. The extended UB and LB are determined by the following formulas.

$$\begin{cases} \mathrm{UB} = x_{\max} + \mathrm{IT} \times \frac{R}{n} \\ \mathrm{LB} = x_{\min} - \mathrm{DT} \times \frac{R}{n} \end{cases} \tag{5}$$

7. Use the CT, UB, and LB to form a triangular LI function, as shown in Fig. 1. Here we set the LI value of the CT to 1 and the LI values of the boundaries (UB and LB) to 0. Therefore, we can attain the LI values of the existing data through properties of similar triangles.