Energy 84 (2015) 419–431

Contents lists available at ScienceDirect

Energy

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

Mid-term interval load forecasting using multi-output support vector regression with a memetic algorithm for feature selection

Zhongyi Hu [a], Yukun Bao [a, *], Raymond Chiong [b], Tao Xiong [a]

[a] Center for Modern Information Management, School of Management, Huazhong University of Science and Technology, Wuhan 430074, PR China
[b] School of Design, Communication and Information Technology, Faculty of Science and Information Technology, The University of Newcastle, Callaghan, NSW 2308, Australia

ARTICLE INFO

Article history:
Received 29 August 2014
Received in revised form
2 February 2015
Accepted 2 March 2015
Available online 10 April 2015

Keywords:
Interval load forecasting
Multi-output support vector regression
Feature selection
Memetic algorithms
Firefly algorithm

ABSTRACT

Accurate forecasting of mid-term electricity load is an important issue for power system planning and operation. Instead of point load forecasting, this study aims to model and forecast mid-term interval loads up to one month in the form of interval-valued series consisting of both peak and valley points by using MSVR (Multi-output Support Vector Regression). In addition, an MA (Memetic Algorithm) based on the firefly algorithm is used to select proper input features among the feature candidates, which include time lagged loads as well as temperatures. The capability of this proposed interval load modeling and forecasting framework to predict daily interval electricity demands is tested through simulation experiments using real-world data from North America and Australia. Quantitative and comprehensive assessments are performed and the experimental results show that the proposed MSVR-MA forecasting framework may be a promising alternative for interval load forecasting.

1. Introduction

Accurate prediction of mid-term electricity load plays a vital role in making important decisions on the operation and planning of a power utility system. It is useful for maintenance scheduling, power generation expansion and security assessment [1–3]. It also provides a fundamental piece of information used for negotiations of bilateral contracts and development of cost efficient fuel purchasing strategies [3,4]. Recognizing the importance of mid-term load forecasting, a wide variety of studies have developed models ranging from linear ones such as ARIMA (Autoregressive Integrated Moving Average) models [5], exponential smoothing models [6], and regression models [7–9], to nonlinear ones based on NNs (Neural Networks) [1,4,10,11] and SVMs (Support Vector Machines) [7,12,13] for addressing the problem. Reviews of different methodologies and techniques employed for electricity load forecasting can be found in Refs. [14,15].

However, an important point to note from the past studies is their preoccupation with point load forecasting over interval load forecasting. Interval-valued load series, which includes peak loads and valley loads, is an alternative representation to classic load series. As a kind of interval-valued time series, interval-valued loads have the advantage of taking into account variability and/or uncertainty, thus reducing the amount of random variation relative to that found in classic point series [16]. García-Ascanio and Maté [17] pointed out that interval load forecasting provides an important risk management tool when making power system planning and operational decisions in the electricity industry, and it would be interesting to develop interval approaches in practical applications. Besides that, the estimate of future interval load can be useful for load flow analysis [18,19] and transmission network expansion planning [20]. Motivated by the aforementioned advantages of interval load forecasting in practice, this study focuses on developing a machine learning model for daily interval load forecasting for up to one month ahead.

To date, a limited number of interval forecasting models can be found in the relevant literature of various fields (e.g., stock price forecasting [21–23], crude oil price forecasting [24], and temperature forecasting [25]). These models rely on methods such as exponential smoothing [16,25], k-nearest neighbors [21], vector autoregressive [17], vector error correction [22,23], and iMLP (interval Multi-Layer Perceptron) [17,26]. A review and comparison of

* Corresponding author. Tel.: +86 27 87558579; fax: +86 27 87556437.
E-mail addresses: yukunbao@hust.edu.cn, y.bao@ieee.org (Y. Bao).

http://dx.doi.org/10.1016/j.energy.2015.03.054
0360-5442/

some existing interval forecasting techniques was reported in Ref. [21]. Apart from the fact that only a few previous studies have focused on interval forecasting of electricity load (e.g., see Refs. [17,27,28]), most of them did not consider the possible interrelations between peak points and valley points [21]. This work aims to fill the gap.

Recently, MSVR (Multi-output Support Vector Regression), a generalized multi-dimensional SVM that uses a cost function with a hyperspherical intensive zone [29], has been successfully applied to many different areas, such as remote sensing biophysical parameter estimation [30], nonlinear channel estimation [31] and multi-step time series forecasting [32]. In view of its unique multi-output modeling structure, which is capable of modeling the interrelations among output variables, in this paper we use MSVR to model and forecast peak and valley loads simultaneously, taking into consideration the potential underlying dynamics of peak load points and valley load points during predefined time horizons (e.g., at daily level).

Furthermore, we propose an MA (Memetic Algorithm) based on the firefly algorithm (see Section 3 for details) to select input features for the MSVR interval load forecasting framework. Although time lagged loads have been widely used as input features in time series modeling, it is well recognized that daily peak loads and valley loads are sensitive to weather variables [28], especially for temperatures. To include both exogenous variables such as temperatures and time lagged loads, the computational complexity and burden will increase substantially and this constantly leads to a bigger pool of feature candidates. Therefore, a metaheuristic approach such as the MA proposed is required for the feature selection process.

To investigate the performance of this proposed interval load forecasting framework, simulation experiments using daily interval load data from a North-American electric utility and the Australian Energy Market Operator are carried out. We compare our MSVR-MA model with several other well-established forecasting models, and the experimental results show that MSVR is able to outperform the others in the majority of cases tested. Statistical analysis confirms that differences between the results obtained are significant.

The main contributions of this study can be summarized as follows:

1. As aforementioned, although interval load forecasting plays an important role in electrical planning and decision making, only a few studies (e.g., [17]) focusing on interval load forecasting exist in the relevant literature. This study proposes a new alternative for daily interval load forecasting (up to a month ahead).
2. In light of the unique multi-output modeling structure of MSVR, this study extends it to interval load forecasting and confirms its superiority over other models in this domain.
3. Extended from our previous work on interval stock index forecasting using MSVR [33], feature selection based on the MA is proposed to overcome the increased computational cost; and thus the proposed MSVR-MA framework can deal with forecasting tasks beyond pure time series forecasting.
4. By using data from two real-world cases, this study confirms the effectiveness of the proposed approach against some well-established counterparts in the literature via simulation experiments, implying that the proposed MSVR-MA framework can be a promising alternative approach for interval load forecasting.

The remainder of this paper is organized as follows. In Section 2, we describe the data series, notations, along with some preliminary analysis, which is necessary for readers to understand the characteristics of interval load series as well as challenges of the modeling task. In Section 3, we present the proposed interval load modeling and forecasting framework, which comprises the MSVR-based forecasting model and MA-based feature selection approach. Our experimental design and setup are discussed in Section 4, followed by the simulation and statistical test results in Section 5. Finally, we conclude this study in Section 6.

Table 1
Samples of interval-valued load series.

Year 1990	Interval-valued load [Peak, Valley] (MW)
Jan. 01	[1672, 3008]
Jan. 02	[1880, 3349]
Jan. 03	[1840, 3230]
Jan. 04	[1744, 3031]
Jan. 07	[1655, 2926]
…	…
…	…

2. Data description and analysis

The aim of this section is to illustrate the general pattern of load series and the relationship between loads and temperatures, showing that the inclusion of temperatures as exogenous variables (features) is necessary. Our analysis in this section will be based on data from a North-American electric utility,[1] which is also one of the two cases used for the simulation experiments carried out in this work.

In this study, interval load data does not come from noise assumptions, but from the aggregation of hourly electrical demand series, which is basically in the form of interval-valued series. The period of the hourly load series for the North-American electric utility case ranges from January 1, 1985 to October 12, 1992. Specifically, given the hourly load series $\{L_h\}_{t=1}^{n} = \{(l_1, l_2, \ldots, l_{24})\}_{t=1}^{n}$, the interval-valued load on day t is defined as $[L]_t = [L_t^V, L_t^P]$, where $L_t^V = \min_t L_h$ and $L_t^P = \max_t L_h$ are the valley load and peak load on day t, respectively. An interval-valued load series, denoted by $\{[L]_t\}_{t=1}^{n} = \{[L_t^V, L_t^P]\}_{t=1}^{n}$, is a chronological sequence of interval-valued load. Table 1 and Fig. 1 show the samples of interval-valued load from January 1 to 31, 1990.

Climate conditions, such as temperature and humidity, have always been an important factor in load forecasting [34–39]. Unfortunately, the only climate information available in this study is the hourly temperature, but it still makes sense due to its strong association with electricity consumption. As such, temperatures are taken as the exogenous variables (input features) in this study. It should be noted that the interval of daily temperatures, which consists of the highest and lowest temperatures, is considered. The daily interval-valued temperature, $\{[T]_t\}_{t=1}^{n} = \{[T_t^V, T_t^P]\}_{t=1}^{n}$, is constructed from the hourly temperature as done for daily interval-valued load series. The variation of interval-valued load with respect to the interval-valued temperature is shown in Figs. 2 and 3. Fig. 2 shows the plots of interval-valued load and interval-valued temperature in one figure. Fig. 3 shows the correlation between interval-valued loads and interval-valued temperatures by using an interval-valued Scatter plot.[2] As can be seen from

[1] The dataset is available from <http://www.ee.washington.edu/class/555/el-sharkawi/index_files/Page3404.htm>.

[2] The process of generating Fig. 3 is given in Appendix A.