Electric Power Systems Research 146 (2017) 270–285

Contents lists available at ScienceDirect

Electric Power Systems Research

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

Short-term electric load forecasting based on singular spectrum analysis and support vector machine optimized by Cuckoo search algorithm

Xiaobo Zhang[a], Jianzhou Wang[a,*], Kequan Zhang[b]

[a] School of Statistics, Dongbei University of Finance and Economics, Dalian 116025, China
[b] Key Laboratory of Arid Climatic Change and Reducing Disaster of Gansu Province, College of Atmospheric Sciences, Lanzhou University, Lanzhou 730000, China

ARTICLE INFO

Article history:
Received 12 May 2016
Received in revised form
19 December 2016
Accepted 29 January 2017

Keywords:
Singular spectrum analysis
Cuckoo search algorithm
Support vector machine
Electric load forecasting
Forecasting validity

ABSTRACT

Short-term electric load forecasting (STLF) has been one of the most active areas of research because of its vital role in planning and operation of power systems. Additionally, intelligent methods are increasingly popular in forecasting model applications. However, the observed data set is often contaminated and nonlinear by as a result of such that it becomes difficult to enhance the accuracy of STLF. Therefore, the novel model (CS-SSA-SVM) for electric load forecasting in this paper was successfully proposed by the combination of SSA (singular spectrum analysis), SVM (support vector machine) and CS (Cuckoo search) algorithms. First, the signal filtering technique (SSA) is applied for data pre-processing and the novel model subsequently models the resultant series with different forecasting strategies using SVM optimized by the CS algorithm. Finally, experiments of electric load forecasting are used as illustrative examples to evaluate the performance of the developed model. The empirical results demonstrated that the proposed model (CS-SSA-SVM) can improve the performance of electric load forecasting considerably in comparison with other methods (SVM, CS-SVM, SSA-SVM, SARIMA and BPNN).

1. Introduction

Over the past few decades, accurate electric load forecasting has gained increasing attention since it plays a significant role in the secure and economic operation of power systems. However, the accuracy of electric load forecasting cannot often fulfill our desired result because it is influenced by various uncertain and uncontrollable factors, such as climate change, economic development, human social activities and country policies. Consequently, it is difficult for us to improve the accuracy of load forecasting. Taking into account all the factors is hardly realistic for the forecasting model. Given this, improving the accuracy of load forecasting is feasible by developing a method to consider key factors.

Over the past few decades, numerous methods have been applied to compute an accurate load forecasting, such as some traditional classical prediction models including regression methods [1–3], exponential smoothing [4–7], ARMA model [8], ARIMA model [9,4,10], season ARIMA [11], grey forecasting model (GM) [12,13], Kalman filter [14], etc.; all of these methods can achieve electric load forecasting but cannot receive the desired prediction accuracy because of their limitations. For example, linear regression depends on historical data and cannot solve non-linear problems. Autoregressive moving average models give the result taking only into account the past and current data points while ignoring other influential elements. The grey forecasting model can only effectively solve the problem with exponential growth trends. To overcome limitations, in recent years, more effective methods have been increasingly proposed to forecast electric loads, such as artificial neural network (ANN) [15–17], feed-forward multilayer perceptron (MLP), radial basis function (RBF), fuzzy logic [18–20] and expert system [21–23]. Though they outperform traditional methods, they cannot reach the satisfactory accuracy because of their defects. For example, artificial neural networks are subject to fall into a local minimum and expert systems strongly depend on knowledge databases. This leads to the rapid development of integrated and hybrid models with a combination of different individual models. For instance, Li et al. [24] proposed an integrated model by combining the generalized regression neural network and

* Corresponding author at: School of Statistics, Dongbei University of Finance and Economics, Dalian 116025, China. Tel.: +86 13130476268; fax: +86 41184710484.
E-mail address: wangjz@dufe.edu.cn (J. Wang).

http://dx.doi.org/10.1016/j.epsr.2017.01.035

the fruit fly optimization algorithm. Hong [25] applied the CPSO algorithm to determine the parameters of SVR. However, Che and Wang [26] developed a hybrid model called SVRARIMA that sums up the SVR and ARIMA models. Valenzuela et al. [27] developed a hybrid intelligent model that sums up fuzzy systems, evolutionary algorithms, and ANNs. Liu et al. [28] presented a hybrid model with parameter optimization that sums up extended Kalman filter, extreme learning machine, EMD and PSO. Additionally, all of the models above had better forecasting performance than individual model.

It can be observed that, through the analysis of the literature over the past few years, the combination of integrating different approaches has become an increasing trend. In this regard, intelligent algorithms applied for optimizing parameters seem to have been a fundamental and popular topic because their global searching capability can overcome the limitation of artificially selecting parameters of forecasting models to ensure the improvement of forecasting performance. In addition, because of the inherent characteristics of original data, for example, the chaotic features caused by noise signal, the noise signal filtering technique utilized for data processing also becomes a crucial data pre-processing stage. Hence, the hybrid or combination model can obtain better forecasting performance than the individual model. In this sense, there are many available forecasting methods, optimization algorithms and data processing techniques for developing different hybrid models. However, there are no acknowledged references to select different methods to build a hybrid model.

As previously mentioned, an artificial neural network is widely used as the forecasting model, but the neural network easily falls into the local minimum because of the restriction on generalization ability and cannot make full use of information from selecting sample with a small sample size [29]. Compared with a traditional neural network, the support vector machine (SVM) can overcome these drawbacks to improve forecasting performance. As a kernel-based method, SVM employs the learning principle with structural risk minimization (SRM) to increase its generalization capability in the training process, generating better forecasts. Because of the attractive feature and empirical performance, SVM has become one of the most promising and popular forecasting methods [30]. Therefore, SVM in this paper is used as the forecasting method. However, the parameters of SVM have an important influence on the accuracy of prediction [31]. An alternative to solve the problem is by using heuristic optimization algorithms for parameter selection that they are prone to be more efficient and robust than a traditional optimization algorithm, e.g., grid search algorithm. Therefore, the Cuckoo search (CS) algorithm, a heuristic optimization algorithm that has powerful ability to search for an optimal solution, is used to determine parameters of SVM. In addition, Singular Spectrum Analysis (SSA), a powerful technique in time series analysis [32] that was used for electric load forecasting [33], is employed to remove the high frequency components of the noisy load series in order to improve the forecasting performance of the SVM model, producing the CS-SSA-SVM model.

In this work, a novel hybrid model is proposed by the combination of the data preprocessing-based technique, kernel-based method and heuristic optimization algorithm. The performance of the hybrid model will be validated by forecasting the short-term electric load. The major contributions of this paper are given as follows:

(1) This paper proposes a novel hybrid forecast model that can integrate merits of individual algorithms to enhance prediction accuracy. In the proposed model, based on the signal preprocessing technique, the proposed hybrid method can robustly map input space into feature space to tackle the complex nonlinear problems.
(2) The powerful signal processing technique is applied to decomposition and reconstruction of the original electric load data; the analysis process from embedding to diagonal averaging performs identification and extraction of different characteristics to alleviate negative effects from the noisy signal.
(3) The powerful global search capacity of the CS algorithm is employed to serve for optimal intelligent selection of model parameters, overcoming limitations of artificially selected parameters.
(4) Apply the proposed hybrid model to the constructed electric load series to show its superiority compared with other benchmark models.

The rest of this paper is organized as follows: Section 2 introduces the Singular Spectrum Analysis technique (SSA). Section 3 introduces the Cuckoo search (CS) algorithm. Section 4 presents the support vector machine (SVM) theory. Section 5 presents the framework of the proposed hybrid forecasting model. Section 6 describes the experimental setup and the results of SSA. Section 8 gives the forecasting results and comparisons among models. Finally, the study's conclusions are presented in Section 9.

List of abbreviations:

SSA Singular Spectrum Analysis
CS Cuckoo Search
SVM Support Vector Machine
SSA-SVM The combination of Singular Spectrum Analysis and Support Vector Machine
CS-SVM The combination of Cuckoo Search and Support Vector Machine
CS-SSA-SVM The combination Cuckoo Search, Singular Spectrum Analysis and Support Vector Machine
MAPE Mean Absolute Percent Error
MAE Mean Absolute Error
MSE Mean Square Error
BPNN BP Neural Network
ARIMA Autoregressive Integrated Moving Average
SARIMA Seasonal Autoregressive Integrated Moving Average

2. Singular Spectrum Analysis (SSA)

SSA, a powerful and reliable technique for time series analysis, is mainly used to identify and extract trends, and periodic or quasi-periodic components or noise. As a relative novel nonparametric technique, SSA performs time series analysis incorporating the elements of classical time series analysis, multivariate statistics, geometry, dynamical systems and signal processing, making it successful for application in biology, physics, climatology, and economics [34–37]. However, the technique cannot be widely used for time series forecast.

2.1. Basic SSA

Generally, SSA can be implemented by two stages: decomposition and reconstruction of basic SSA. The first stage includes embedding and singular value decomposition (SVD). The reconstruction stage consists of grouping and diagonal averaging. Therefore, a SSA is performed by four steps: embedding, singular values decomposition (SVD), grouping and diagonal averaging [38]. A mathematical review of SSA is described in the following.

Step 1: Embedding. Consider time series $\boldsymbol{Y}=[y_1, y_2, \ldots, y_N]^T$, available at N point. The dimension of embedding in the SSA