Energy 113 (2016) 796–808

Contents lists available at ScienceDirect

Energy

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

An incremental electric load forecasting model based on support vector regression

YouLong Yang [a], JinXing Che [a,*], YanYing Li [b], YanJun Zhao [c], SuLing Zhu [d]

[a] School of Mathematics and Statistics, Xidian University, 266 Xinglong Section of Xifeng Road, Xi'an, Shaanxi 710126, China
[b] College of Mathematics and Information Science, Baoji University of Arts and Sciences, Baoji 721013, Shaanxi, China
[c] Humanities and Sciences College, Northeast Normal University, Changchun 130117, China
[d] School of Public Health, Lanzhou University, Lanzhou 730000, China

ARTICLE INFO

Article history:
Received 30 January 2016
Received in revised form
29 June 2016
Accepted 19 July 2016

Keywords:
Electric load forecasting
Phase space reconstruction
Support vector regression
Representative data set reconstruction method
Nested particle swarm optimization

ABSTRACT

With the smart portable systems and the daily growth of databases on the web, there are ever-increasing requirements to learn the batch arriving and large sample data set. In this paper, an incremental learning model of support vector regression (SVR) is proposed to forecast the electric load under the batch arriving and large sample. For modeling with SVR, the optimal embedding of time series is constructed by phase space reconstruction (PSR). Then, an optimal training subset for the training of SVR is extracted based on the current data set, which enables us to cut the high time and space complexity by reducing the full training data set. When newly-increased data are added into the system, a representative data set reconstruction method is presented for quickly re-training the current SVR, and a nested particle swarm optimization (NPSO) framework is presented to select the parameters of the incremental SVR model. Experiments of incremental electric load forecasting demonstrate the computational superiority of the presented model over the comparison models.

1. Introduction

As electricity is an energy source that cannot be stored in large amounts, accurate modeling of electric load is the cornerstone of the optimal scheduling of electrical production, distribution and consumption [1,2]. Recent years, with the development of the information technology, batch arriving data are input into the power system in real time, which makes the incremental electric load forecasting model become more and more important in modern electricity markets.

As one of the most promising and popular family of machine learning, support vector regression (SVR) has many attractive features and profound empirical performances in the applications of small sample, nonlinearity and high dimensional dataset [3–6]. With the smart portable systems and the daily growth of databases on the web, there are ever-increasing requirements to learn the batch arriving data and large sample data set [7]. However, SVR has the following weaknesses for these data sets: the incomprehensible black-box modeling process, high time complexity $O(N^3)$, and high space complexity $O(N^2)$ (N is the size of training dataset) [8]. To this end, this paper aims to extend the excellent properties of SVR to the incremental situation of batch arriving data: that is, iteratively update the representative data set, and quickly re-train the current SVR whenever new data is obtained.

Only the training data points near decision boundary, called support vectors, have impact on the final prediction model of support vector machine (SVM). Inspired by that, a new type of SVM model extracts support vectors (SVs) to speed up the large sample learning. Nahla H. Barakat and Andrew P. Bradley extract rules directly from the support vectors (SVs) of a trained SVM using SQRex-SVM algorithm, which demonstrates that the SVs provide the possibility of data interpretability [9]. Chang-Dong Wang et al. present a novel data stream clustering algorithm by using support vector domain description and support vector clustering [10]. Jingnian Chen et al. propose an instance selection method especially for multi-class problems in order to speed up SVM training process [8]. These studies demonstrate that the obtained SVM model is determined by the SV set (small part of the full training data set) completely, so the SV set is the most informative data and can represent the full training data set.

However, for the batch arriving data, usual batch learning

* Corresponding author.
E-mail address: jinxingche@163.com (J. Che).

http://dx.doi.org/10.1016/j.energy.2016.07.092
0360-5442/

Nomenclature

y_t	The time series of electric load in time period t
U,V	The linear factors of data preprocessing
τ	The time delay of phase space reconstruction
s	The embedding dimension of phase space reconstruction
T	The length of finite time series
$C(s,T,r,t)$	The correlation integral of C–C method
x_t	The multi-variate vector of phase space reconstruction in time period t
K	The kernel function of SVR
$delta$	The width parameter of kernel function
f	The regression function of SVR
ω,b	The weight vector and the bias of regression function
ε	The maximum value of tolerable error
ξ_i, ξ_i^*	The distance between actual values and the corresponding boundary values of ε-tube
C	The trade-off parameter between generalization ability and training error
g	The fitness evaluation function of PSO
$x_i(h)$	The position of particle i at the moment h
$v_i(h)$	The velocity of particle i at the moment h
p_i,p_g	The itself and the entire swarm best position
c_1,c_2	The weight factors
$\omega(max)$	The maximum inertia weight
$\omega(end)$	The final inertia weight
$MaxIteration$	The maximum iterations
$P_{best}(1)$	The best particle of the original selection process for training set A_V
λ_i	The contraction weight of the nested PSO

schemes use the past data together with the new data to perform training whenever new data are received, thus consuming a lot of time. Lee presents a probabilistic-entropy-based neural network (PENN) model for tackling incremental data regression problems [11]. Rüping et al. introduce incremental learning for SVM firstly and point out that the SV set is the minimum set of the training data set [12]. Recently, incremental SVMs for incremental classification problems have been proposed in many studies, such as [13] and [14].

Nevertheless, few works tackle the issue of incremental learning of support vector regression (SVR). Objective of learning from data is not that the obtained model performs well in the training phase. Rather, a good model should have strong generalization capability and be easy to understand [15]. SVR extracts an appropriately sparse representations (the so-called support vectors) through kernel modeling technique, which has become a popular machine learning model due to its attractive properties and excellent experimental results [16,17]. Inspired by that the farther points contain more information than the nearer ones, Che et al. presents an optimal training subset for support vector regression (SVR) load forecasting model, and determines the size of the optimal training subset by using an approximation convexity optimization algorithm [18]. To solve the parameters selection and slow learning for large sample in most existing SVR learning processes, Che considers an adaptive particle swarm optimization (APSO) algorithm for the parameters selection of the SVR model in Ref. [19], and demonstrates the effectiveness and generalization of the proposed SVR by an UCI data set and an electric load data set [19]. However, an important question arises: When new training data have to be included in the current training set, how can one quickly update the optimal training subset without re-executing the algorithm proposed in Ref. [19]? This is the starting point of the proposed incremental learning algorithm.

In this paper, an incremental learning algorithm for SVR is proposed to forecast the electric load under the batch arriving and large sample. From now on this proposed algorithm will be simply named incremental SVR (I-SVR). In a first step, the raw data are normalized over a range, the dynamics of this data is viewed by performing a phase space reconstruction technique, and the delay time and embedding dimension are computed by implementing the C–C method due to its easy operation and low computational complexity. Then, an original SVR model is trained by using an optimal training subset method. When new input data are added into the system, a representative data set reconstruction method is presented for quickly re-training the current SVR, then a nested particle swarm optimization (NPSO) framework is employed for the parameters selection of the incremental SVR model. The proposed representative data set reconstruction method can simplify the batch learning algorithm and increase the learning speed. This proposed algorithm is preferred over batch learning algorithms as it does not require the above retraining whenever new data are received. Experiments of incremental electric load forecasting demonstrate the computational superiority of the presented model over the comparison models.

Our key contributions in this paper are summarized in the following items: i) introduce the problem of incremental electric load forecasting under the SVR framework, ii) propose a representative data reconstruction method to update the "support vector candidates for quickly re-training the current SVR, and iii) develop a nested PSO algorithm to reduce the space search of the parameter selection for I-SVR.

This paper is structured as follows: Section 2 gives a short introduction of initial SVR modeling for electric load forecasting. In Section 3, the proposed incremental SVR for forecasting is described in detail, and the main steps of the model are given. In Section 4, the data description and research design are introduced, next, empirical results obtained and comparisons are presented and discussed. Section 5 concludes this paper with a brief summary and presents the future research.

2. Background

Recently, SVR has become a popular and effective forecast model for forecasting electric load [20–22]. Chia et al. propose a SVR model to accurately forecast the load demand in advance for a solar energy application [23]. Based on SVR and fuzzy-rough feature selection with particle swarm optimization algorithms, Son and Kim propose a forecasting model to forecast the short-term electricity demand in residential sector [24]. By using the Phase Space Reconstruction procedure, Santamaría-Bonfil et al. present a hybrid model based on SVR to forecast the univariate wind speed time series [25].

In this section, we present the initial modeling with SVR, which includes data preprocessing, phase space reconstruction for a unidimensional time series, model building and parameters selection of SVR. Based on this initial modeling, we propose the novel incremental learning algorithm for SVR model in the next Section.