Applied Soft Computing 25 (2014) 15–25

Contents lists available at ScienceDirect

Applied Soft Computing

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

Comprehensive learning particle swarm optimization based memetic algorithm for model selection in short-term load forecasting using support vector regression

Zhongyi Hu, Yukun Bao *, Tao Xiong

Department of Management Science and Information Systems, School of Management, Huazhong University of Science and Technology, Wuhan 430074, PR China

A R T I C L E I N F O

Article history:
Received 18 December 2013
Received in revised form 4 August 2014
Accepted 14 September 2014
Available online 20 September 2014

Keywords:
Support vector regression
Short-term load forecasting
Model selection
Memetic algorithms
Particle swarm optimization

A B S T R A C T

Background: Short-term load forecasting is an important issue that has been widely explored and examined with respect to the operation of power systems and commercial transactions in electricity markets. Of the existing forecasting models, support vector regression (SVR) has attracted much attention. While model selection, including feature selection and parameter optimization, plays an important role in short-term load forecasting using SVR, most previous studies have considered feature selection and parameter optimization as two separate tasks, which is detrimental to prediction performance.
Objective: By evolving feature selection and parameter optimization simultaneously, the main aims of this study are to make practitioners aware of the benefits of applying unified model selection in STLF using SVR and to provide one solution for model selection in the framework of memetic algorithm (MA).
Methods: This study proposes a comprehensive learning particle swarm optimization (CLPSO)-based memetic algorithm (CLPSO-MA) that evolves feature selection and parameter optimization simultaneously. In the proposed CLPSO-MA algorithm, CLPSO is applied to explore the solution space, while a problem-specific local search is proposed for conducting individual learning, thereby enhancing the exploitation of CLPSO.
Results: Compared with other well-established counterparts, benefits of the proposed unified model selection problem and the proposed CLPSO-MA for model selection are verified using two real-world electricity load datasets, which indicates the SVR equipped with CLPSO-MA can be a promising alternative for short-term load forecasting.

1. Introduction

Short-term load forecasting (STLF) aims to predict electricity loads over a short time period. Traditionally, it has been considered a very important issue since not only is it critical for automatic generation control, reliable operation, and resource dispatch, but it also contains fundamental information used for energy transactions in competitive electricity markets [1,2]. However, the electricity load is inevitably affected by various factors, such as climate, social activities, and seasonal factors, and the prediction performance is also highly dependent on the modeling configuration, such as parameter tuning for a specific model. Therefore, it is very difficult to forecast the electricity load accurately when faced with challenges arising from the vast selection of candidate input variables and model configuration parameters.

During the past few decades, many approaches for load forecasting have been proposed, such as the autoregressive moving average model [3], regression models [4,5], expert systems [6], fuzzy logic [7], semi-parametric additive model [8], functional time-series predictor [9], neural networks (NN) [2,10–12], and support vector machines (SVMs) [13–20]. Among these, support vector regression (SVR, a regression form of SVMs) is a powerful machine learning technique with strong theoretical foundation [21,22] and has been obtained appealing performance in the field of load forecasting, for example [13–19]. As [23] advocated, the electrical load forecasting has been one of the most widely studied applications for SVR and its variations (i.e., least squares SVMs). Aimed at improving modeling quality, previous research efforts have mainly focused on selecting

* 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.asoc.2014.09.007

the influencing input variables (feature selection) and tuning the appropriate parameter settings of the SVR model.

Many factors, such as historical load, meteorological factors, and calendar information, have been examined in previous studies [1,11,15,18,24–26]. Of these influencing factors, some could be redundant or even irrelevant to a specific STLF problem; yet there is no general rule about which influencing factors and especially how many time lags of the factors should be included in the case of a time series forecasting system. This is a strong motivation for a **feature selection technique** in the field of STLF problem [18,27]. Feature selection techniques used in past studies can be divided into two types: filter methods and wrapper methods. A filter method chooses the feature subset based on an evaluation criterion such as mutual information (MI) [28–31], Bayesian 'automatic relevance determination' [32], or correlation and linear independency [33,34], maximum-relevance minimum-redundancy criterion (MRMR) and ReliefF [18]. The filter method is characterized by its independence of the learning algorithm and mainly focuses on the invention of measures depicting the relationship between each subset of input variables and the output. As opposed to a filter method, wrapper methods use the performance of the forecasting model as an evaluation criterion to identify the correct input subset. Various wrapper methods have been investigated, and the implementation of some metaheuristics such as the simulated rebounding algorithm [24], simulated annealing [27], genetic algorithms and ant colony optimization [35] are regarded as common practice.

Parameter optimization is concerned with the optimal setting of parameters in SVR, such as the penalty coefficient, the kernel parameters, and the width of the loss function. The selection of these parameters is crucial to obtain good performance in handling the electricity load forecasting task using SVR. By involving metaheuristics, such as GAs [36,37], chaotic particle swarm optimization [38], and artificial bee colony algorithms [39], particle swarm pattern search method [18], simulated annealing [27], various studies have focused particularly on parameter optimization in an SVR forecasting model for electricity loads.

However, the main disadvantage of the above studies with regard to feature selection and parameter optimization is that they address these two subtasks in an almost disjoint way. To understand this consideration, we must bear in mind that although parameter optimization and feature selection are two separate issues in model selection, the feature subset choice influences the appropriate parameters, and vice versa [40]. On the contrary, this study investigates model selection with a lens on the development of a unifying modeling framework. Obviously, a unifying implementation of these two subtasks implies increased complexity in the problem. As a powerful algorithmic paradigm for evolutionary computing in a wide variety of areas [41–48], MAs are especially attractive for model selection because of their powerful search ability, both in exploration and exploitation. Therefore, this study proposed a solution to the proposed model selection problem in the framework of MAs. Specifically, considering the successful performance in existing applications [49–53], comprehensive learning particle swarm optimization (CLPSO) is applied as global searcher in the proposed MA for exploration of the search space and detection of regions that can potentially yield an optimum solution. Besides, a problem-specific local search is proposed to effectively exploit the potential regions identified by CLPSO in the proposed CLPSO-based MA (CLPSO-MA). The performance of the proposed CLPSO-MA based model selection in SVR is compared using real-world electricity loads with certain selected established counterparts.

The main aims of this study are to make practitioners aware of the benefits of applying model selection in STLF using SVR and to provide one solution for model selection using the proposed CLPSO-MA algorithm. The contributions of this work can be summarized as follows. (a) This study first considers feature selection and parameter optimization as a unified model selection process in STLF using SVR. Although many studies have paid special attention on the feature selection and parameters optimization in the SVR-based load forecasting modeling, few, if any, studies related to unified model selection for load forecasting using SVR have been reported in the literature. (b) A novel memetic algorithm is proposed to solve the model selection task; this could be considered to be the first application of an MA for model selection in the case of STLF. (c) In the proposed MA, CLPSO is applied to explore the search space and detect the regions that can potentially yield an optimum solution. A special local search is also proposed to address the local search of the unified model selection problem which is essentially a continuous-binary optimization problem. (d) Compared with other well-established counterparts, benefits of the proposed unified model selection problem and the proposed CLPSO-MA for model selection are verified using two real-world electricity load datasets, which indicates the SVR equipped with CLPSO-MA can be a promising alternative for short-term load forecasting.

The remainder of this paper is organized as follows. Section 2 discusses the basic concept and technical concerns of SVR modeling in STLF. Section 3 elaborates on the proposed CLPSO-MA based model selection for SVR forecasting model. Section 4 shows the details of experimental setup. The results together with a discussion thereof are reported in Section 5. Finally, we conclude the paper in Section 6.

2. Modeling with support vector regression

2.1. Formulation

The essence of STLF is a type of regression procedure. Consider the pretreated sample set $\{(x_i, y_i)\}_{i=1}^{n}$, where $x_i \in \mathbf{R}^m$ is the i-th vector containing m input features, y_i is the corresponding desired load, and n denotes the number of data items in the sample set. Based on Vapnik's statistical learning theory [22], SVR generates a mapping function using the optimization problem:

$$\begin{aligned}
&\text{Minimize} && R = C\sum_{i=1}^{n}(\xi_i+\xi_i^*) + \frac{1}{2}\parallel w\parallel^2 \\
&\text{Subject to} && y_i - < w, \phi(x_i) > -b \le \varepsilon + \xi_i \\
& && < w, \phi(x_i) > +b - y_i \le \varepsilon + \xi_i^* \\
& && \xi_i, \xi_i^* \ge 0, \quad i = 1, 2, \ldots, n
\end{aligned} \tag{1}$$

where C is a penalty parameter. Here ξ_i and ξ_i^* are non-negative slack variables, and $\phi(x)$ is the high-dimensional feature space, which is non-linearly mapped from the input space x.

According to Wolfe's Dual Theorem and the saddle-point condition, the dual optimization problem of the above problem is obtained as follows:

$$\begin{aligned}
\max_{\alpha,\alpha^*} \quad & -\frac{1}{2}\sum_{i,j=1}^{l}(\alpha_i-\alpha_i^*)(\alpha_j-\alpha_j^*) < \phi(x_i), \phi(x_j) > \\
& -\varepsilon\sum_{i,j=1}^{l}(\alpha_i+\alpha_i^*) + \sum_{i,j=1}^{l} y_i(\alpha_i-\alpha_i^*) \\
s.t. \quad & \sum_{i,j=1}^{l}(\alpha_i-\alpha_i^*) = 0 \quad \text{and} \quad \alpha_i, \alpha_i^* \in [0, C]
\end{aligned} \tag{2}$$