Applied Energy 132 (2014) 602–609

Contents lists available at ScienceDirect

Applied Energy

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

Short-term load forecasting using a kernel-based support vector regression combination model

JinXing Che [a,*], JianZhou Wang [b]

[a] College of Science, NanChang Institute of Technology, NanChang 330099, JiangXi, China
[b] School of Mathematics and Statistics, LanZhou University, LanZhou, GanSu 730000, China

HIGHLIGHTS

- Propose a kernel-based support vector regression combination model.
- Combine models by using a novel individual model selection algorithm.
- Provide a new way to kernel function selection of SVR model.
- The performance and electric load forecast accuracy are assessed by two real cases.
- Experiments show the superiority of the combination model compared to single kernel.

ARTICLE INFO

Article history:
Received 17 February 2014
Received in revised form 10 June 2014
Accepted 17 July 2014

Keywords:
Short-term load forecasting
Kernel
Support vector regression
Combination model
Selection algorithm

ABSTRACT

Kernel-based methods, such as support vector regression (SVR), have demonstrated satisfactory performance in short-term load forecasting (STLF) application. However, the good performance of kernel-based method depends on the selection of an appropriate kernel function that fits the learning target, unsuitable kernel function or hyper-parameters setting may lead to significantly poor performance. To get the optimal kernel function of STLF problem, this paper proposes a kernel-based SVR combination model by using a novel individual model selection algorithm. Moreover, the proposed combination model provides a new way to kernel function selection of SVR model. The performance and electric load forecast accuracy of the proposed model are assessed by means of real data from the Australia and California Power Grid, respectively. The simulation results from numerical tables and figures show that the proposed combination model increases electric load forecasting accuracy compared to the best individual kernel-based SVR model.

1. Introduction

The short-term load forecasting (STLF) has a definitive impact on the daily operations of a power utility [1]. It is used for various purposes, such as price and income elasticities, energy transfer scheduling, unit commitment and load dispatch. With the emergence of load management strategies, the load prediction has played a broader role in utility operations [2,3]. Thus, the development of an accurate, fast, simple and robust load prediction algorithm is important to electric utilities and its customers.

Kernel-based methods, such as support vector machines (SVMs) and Gaussian processes, have became one of the most promising and popular family of learners due to their attractive features and profound empirical performance in a wide variety of supervised and non-supervised learning tasks [4,5]. The main steps in the design of these learning algorithms are as follows: the first is to map the training data x from the input space χ into some other (usually higher dimensional) feature space F, and the second is to apply a linear procedure in F. The major effort in kernel-based methods is the selection of an appropriate kernel function, that is, the kernel function somehow fits the learning target. Unsuitable kernel function or hyper-parameters setting may lead to significantly poor performance [6].

However, the researchers need to select in advance the type of kernel function and the associated kernel hyper-parameters for SVM [7]. Consider that the Gaussian kernel has strong generalization capability, the existing studies on using SVM are very limited in that usually only the Gaussian kernel function and the associated parameters are selected and studied in these works [8]. To compare the performance of different kernels, Zhou et al. implement three SVM kernels, namely linear, Gaussian, and polynomial

* Corresponding author.

http://dx.doi.org/10.1016/j.apenergy.2014.07.064
0306-2619/

kernels, and compute the optimal combination of the associated kernel hyper-parameters by fine tuning, then conclude that the three kernels give comparable forecasting accuracy [9]. Asraf et al. evaluate SVM classifier with three different kernels, namely linear kernel, polynomial kernel with soft margin and polynomial kernel with hard margin, and demonstrate that polynomial kernel with soft margin is capable of classifying nutrient diseases accurately in the oil palm leaves with accuracy of 95% of correct classification [10]. These investigations provide answers to some fundamental questions such as which kernel function should be chosen?

For some complex problems (for example, load forecasting), a single kernel may not be sufficient to describe the data characteristics satisfactorily [11]. To combine the advantages of different kernels, some researchers have adopted multiple-kernels to deal with these problems [12–14]. The idea that combining different kernel functions might be worthwhile has gained wide acceptance since the seminal article of Lanckriet et al. [15]. For this situation, one can obtain a complex kernel by linear combining simper ones, but how can one select the optimal combination subset from all individual kernels, and judge whether the resulting kernel is better or worse than its components? This is the starting point of our study.

Recently, an interesting solution, measuring the degree of agreement between a kernel and a given learning task, has been developed through the concept of kernel-target alignment (KTA) [16]. For a classification task, the ideal kernel for a classification target $y(x)$ is $K(x,z) = y(x)y'(z)$, so the goodness measurement of a kernel K corresponds to the alignment between the K and the ideal kernel yy'. Nguyen and Ho study the problem of evaluating the goodness of a kernel matrix for a classification task, and show that the above kernel target alignment (KTA) has some serious drawbacks, then present a feature space-based kernel matrix evaluation measurement to overcome the limitations of KTA [17]. To deal with kernel fusion problem and give more flexibility to kernel function, multiple kernel learning (MKL) optimization goal, which considers a group of kernels simultaneously, is estabilished by using KTA method [18,19]. Previous works on KTA focused mainly on classification problem by linear combination of kernels in a transductive or inductive settings. For a regression task, the ideal kernel, however, is hard to determine, and has been rarely studied.

To overcome this problem, combination model is firstly employed to combine multiply kernels based on a novel combination selection algorithm. The combination selection algorithm select the optimal subset of individual support vector regression (SVR) kernel models from all available SVR kernel models using the proposed goodness measurement. As indicated above, the proposed combination model solves the above difficulty from a new perspective. Australia and California Power Grid short-term load demands are used as cases study for load forecasting performance testing. The results demonstrate that the proposed kernel-based SVR combination model has better performance than the best single kernel-based SVR STLF approach.

2. The explicit process of the new algorithm

In this section, the author describe the fundamental background of support vector regression (SVR) kernel models and combination forecasting model, then describe the entropy and mutual information, and also propose the selection algorithm used in this study.

2.1. Support vector regression kernel models

Given a training set $\Gamma = \{x_i, y_i\}_{i=1}^{n}$ where $x_i \in \mathbb{R}^d$ and $y_i \in \mathbb{R}$, the aim of SVR is to induce a forecast which has good forecasting performance on future unseen examples. When the data set Γ is linearly dependent, SVR solves the following problem [20].

$$\min_{\omega,b,\xi,\xi^*} \frac{1}{2}\omega^T\omega + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \tag{1}$$

$$s.t.\begin{cases} y_i - (\omega^T x_i + b) \leqslant \varepsilon + \xi_i \\ (\omega^T x_i + b) - y_i \leqslant \varepsilon + \xi_i^*, \\ \xi_i, \xi_i^* \geqslant 0 \end{cases} \tag{2}$$

where ε denotes the maximum value of tolerable error, ξ_i and ξ_i^* is the distance between actual values and the corresponding boundary values of ε-tube, $C > 0$ decides the trade-off of generalization ability and training error. This problem can be solved by making use of the Karush–Kuhn–Tucker's (KKT) conditions [4]. Then, the traditional SVR would be the following linear regression function

$$f(x) = \omega^T x + b, \tag{3}$$

where ω represents the weight vector; b represents the bias.

By employing "kernel trick", SVR has been extended to solve nonlinear regression problems with a linear method in an appropriate feature space [21]. Thus, the performance of SVR is determined by the type of kernel function and the settings of kernel parameters.

The following four types of kernel function, namely linear, tanh, polynomial, and Gaussian kernels, are commonly employed in the related area. The linear kernel is

$$K(x,z) = x^T z, \tag{4}$$

the tanh kernel is

$$K(x,z) = tanh(gx^T z + c), \tag{5}$$

the polynomial kernel is

$$K(x,z) = (x^T z + c)^d, \tag{6}$$

and the Gaussian kernel is

$$K(x,z) = exp\left(\frac{-(x-z)^2}{2 \times \delta^2}\right), \tag{7}$$

where g is the slope of the tanh kernel (positive scalar), c is the offset of polynomial and tanh kernel (scalar, negative for tanh), d is the degree of the polynomial kernel (positive scalar), δ is the width of Gauss kernel (positive scalar). In this study, the SVR kernel models are trained by the method proposed in [22].

2.2. Combination forecasting model

Kernel-based methods have been widely used for time series data analysis and forecasting. For example, the SVR modeling approach proposed by Vapnik [20] has been demonstrated to be effective in load forecasting applications. In a practical situation, in applying the SVR modeling approach, however, one faces the important issue of how to choose the best kernel function among a variety of candidates. Generally speaking, single kernel selection is often unstable and may cause an unnecessarily high variability in the final forecasting model. In this work, the author propose the use of a model selection algorithm to convexly combine the selected kernels for a better performance of prediction. In an environment where individual kernels are subject to structural breaks and misspecified by varying degrees, a strategy that pools information from the multiply kernels learning typically performs better than methods that try to select the best individual kernel.

After selecting the individual kernel-based SVR models, the mostly weighted averages of each of the individual forecasts are