Expert Systems with Applications 36 (2009) 4725–4735

Contents lists available at ScienceDirect

Expert Systems with Applications

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

A Novel hybrid genetic algorithm for kernel function and parameter optimization in support vector regression

Chih-Hung Wu [a,*], Gwo-Hshiung Tzeng [b,c], Rong-Ho Lin [d]

[a] Department of Digital Content and Technology, National Taichung University No. 140, Ming-Shen Road, Taichung 40306, Taiwan
[b] Department of Business Administration, Kainan University, No. 1, Kainan Road, Luchn, Taoyuan 338, Taiwan
[c] Institute of Management of Technology, National Chiao Tung University, 100, Ta-Hsueh Road, Hsinchu 300, Taiwan
[d] Department of Industrial Engineering & Management, National Taipei University of Technology, No. 1, Section 3, Chung-Hsiao East Road, Taipei 106, Taiwan, ROC

ARTICLE INFO

Keywords:
Support vector regression (SVR)
Hybrid genetic algorithm (HGA)
Parameter optimization
Kernel function optimization
Electrical load forecasting
Forecasting accuracy

ABSTRACT

This study developed a novel model, HGA-SVR, for type of kernel function and kernel parameter value optimization in support vector regression (SVR), which is then applied to forecast the maximum electrical daily load. A novel hybrid genetic algorithm (HGA) was adapted to search for the optimal type of kernel function and kernel parameter values of SVR to increase the accuracy of SVR. The proposed model was tested at an electricity load forecasting competition announced on the EUNITE network. The results showed that the new HGA-SVR model outperforms the previous models. Specifically, the new HGA-SVR model can successfully identify the optimal type of kernel function and all the optimal values of the parameters of SVR with the lowest prediction error values in electricity load forecasting.

1. Introduction

Support vector machines (SVMs) have been successfully applied to a number of applications such as including handwriting recognition, particle identification (e.g., muons), digital images identification (e.g., face identification), text categorization, bioinformatics (e.g., gene expression), function approximation and regression, and database marketing, and so on. Although SVMs have become more widely employed to forecast time-series data (Tay & Cao, 2001; Cao, 2003; Kim, 2003) and to reconstruct dynamically chaotic systems (Müller et al., 1997; Mukherjee, Osuna, & Girosi, 1997; Mattera & Haykin, 1999; Kulkarni, Jayaraman, & Kulkarni, 2003), a highly effective model can only be built after the parameters of SVMs are carefully determined (Duan, Keerthi, & Poo, 2003).

Min and Lee (2005) stated that the optimal parameter search on SVM plays a crucial role in building a prediction model with high prediction accuracy and stability. The kernel-parameters are the few tunable parameters in SVMs controlling the complexity of the resulting hypothesis (Cristianini, Campell, & Taylor, 1999). Shawkat and Kate (2007) pointed out that selecting the optimal degree of a polynomial kernel is critical to ensure good generalization of the resulting support vector machine model. They proposed an automatic selection for determining the optimal degree of polynomial kernel in SVM by Bayesian and Laplace approximation method estimation and a rule based meta-learning approach. In addition, to construct an efficient SVM model with RBF kernel, two extra parameters: (a) sigma squared and (b) gamma, have to be carefully predetermined. However, few studies have been devoted to optimizing the parameter values of SVMs. Evolutionary algorithms often have to solve optimization problems in the presence of a wide range of problems (Dastidar, Chakrabarti, & Ray, 2005; Shin, Lee, Kim, & Zhang, 2005; Yaochu & Branke, 2005; Zhang, Sun, & Tsang, 2005). In these algorithms, genetic algorithms (GAs) have been widely and successfully applied to various types of optimization problems in recent years (Goldberg, 1989; Fogel, 1994; Cao, 2003; Alba & Dorronsoro, 2005; Aurnhammer & Tonnies, 2005; Venkatraman & Yen, 2005; Hokey, Hyun, & Chang, 2006; Cao & Wu, 1999; McCall, 2005). Therefore, this paper proposes a hybrid genetic-based SVR model, HGA-SVR, which can automatically optimize the SVR parameters integrating the real-valued genetic algorithm (RGA) and integer genetic algorithm, for increasing the predictive accuracy and capability of generalization compared with traditional machine learning models.

In addition, a wide range of approaches including time-varying splines (Harvey & Koopman, 1993), multiple regression models (Ramanathan, Engle, Granger, Vahid-Araghi, & Brace, 1997), judgmental forecasts, artificial neural networks (Hippert & Pedreira, 2001) and SVMs (Chen, Chang, & Lin, 2004; Tian & Noore, 2004) have been employed to forecast electricity load. One of the most crucial demands for the operation activities of power systems is short-term hourly load forecasting and the extension to several days in the future. Improving the accuracy of short-term load forecasting (STLF) is becoming even more significant than before due to the changing structure of the power utility industry (Tian &

* Corresponding author. Tel.: +886 939013100; fax: +886 422183270.
E-mail addresses: chwu@ntcu.edu.tw (C.-H. Wu), ghtzeng@cc.nctu.edu.tw, ghtzeng@mail.knu.edu.tw (G.-H. Tzeng).

0957-4174/$ - see front matter Crown Copyright © 2008 Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2008.06.046

Noore, 2004). SVMs have been applied to STLF and performed well. Unfortunately, there is still no consensus as to the perfect approach to electricity demand forecasting (Taylor & Buizza, 2003).

Several studies have proposed optimization methods which used a genetic algorithm for optimizing the SVR parameter values. To overcome the problem of SVR parameters, a GA-SVR has been proposed in a earlier paper (Hsu, Wu, Chen, & Peng, 2006) to take advantage of the GAs optimization technique. However, few studies have focused on concurrently optimizing the type of SVR kernel function and the parameters of SVR kernel function. The present study proposed a novel and specialized hybrid genetic algorithm for optimizing all the SVR parameters simultaneously. Our proposed method was applied to predicting maximum electrical daily load and its performance was analyzed. An actual case of forecasting maximum electrical daily load is illustrated to show the improvement in predictive accuracy and capability of generalization achieved by our proposed HGA-SVR model.

The remainder of this paper is organized as follows. The research gap for obtaining optimal parameters in SVR is reviewed and discussed in Section 2. Section 3 details the proposed HGA-SVR, ideas and procedures. In Section 4 an experimental example for predicting the electricity load is described to demonstrate the proposed method. Discussions are presented in Section 5 and conclusions are drawn in the final Section.

2. Basic ideas of methods for obtaining optimal parameters in SVR

SVR is a promising technique for data classification and regression (Vapnik, 1998). We briefly introduce the basic idea of SVR in the Section 2.1. To design an effective model, the values of the essential parameters in SVR must be chosen carefully in advance (Duan et al., 2003). Thus, various approaches to determine these values are discussed in Section 2.2. Although many optimization methods have been proposed, GAs is well suited to the concurrent manipulation of models with varying resolutions and structures since they can search non-linear solution spaces without requiring gradient information or a priori knowledge of model characteristics (McCall & Petrovski, 1999). The genetic algorithm employed in this study to search for the optimal values of the SVR parameter is illustrated in Section 2.3.

2.1. Support vector regression (SVR)

This subsection briefly introduces support vector regression (SVR), which can be used for time-series forecasting. Given training data $(x_1,y_1),\ldots,(x_l,y_l)$, where x_i are the input vectors and $\boldsymbol{y}_i$ are the associated output values of $\boldsymbol{x}_i$, the support vector regression is an optimization problem:

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

$$\text{Subject to } y_i - (\boldsymbol{\omega}^{\mathrm{T}}\phi(x_i) + b) \leqslant \varepsilon + \xi_i, \tag{2}$$

$$(\boldsymbol{\omega}^{\mathrm{T}}\phi(x_i) + b) - y_i \leqslant \varepsilon + \xi_i^*, \tag{3}$$

$$\xi_i, \xi_i^* \geqslant 0, i = 1, \ldots, l, \tag{4}$$

where l denotes the number of samples, $\boldsymbol{x}_i$ vector of i-sample is dataset mapped to a higher dimensional space by the kernel function ϕ, vector, ξ_i represents the upper training error, and ξ_i^* is the lower training error subject to ε-insensitive tube $|y - (\boldsymbol{\omega}^T\phi(x) + b)| \leqslant \varepsilon$. Three parameters determine the SVR quality: error cost C, width of tube, and mapping function (also called kernel function). The basic idea in SVR is to map the dataset x_i into a high-dimensional feature space via non-linear mapping. Kernel functions perform non-linear mapping between the input space and a feature space. The approximating feature map for the Mercer kernel performs non-linear mapping. In machine learning theories, the popular kernel functions are

$$\text{Gaussian(RBF) kernel}: k(x_i,x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right). \tag{5}$$

$$\text{Polynomial kernel}: k(x_i,x_j) = (1 + x_i \bullet x_j)^d. \tag{6}$$

$$\text{Linear kernel}: k(x_i,x_j) = x_i^{\mathrm{T}}x_j. \tag{7}$$

In Eq. (5), x_i and x_j are input vector spaces; and V denotes the variance-covariance matrix of the Gaussian kernel.

2.2. Parameter optimization

As mentioned earlier, when designing an effective model, values of the two essential parameters in SVR have to be chosen carefully in advance (Duan et al., 2003). These parameters include (1) regularization parameter C, which determines the tradeoff cost between minimizing the training error and minimizing model complexity; and (2) parameter sigma (or d) of the kernel function, which defines the non-linear mapping from the input space to some high-dimensional feature space. This investigation considers only the Gaussian kernel, namely sigma square (V), which is the variance-covariance matrix of the kernel function. Generally speaking, model selection by SVM is still performed in the standard way: by learning different SVMs and testing them on a validation set to determine the optimal value of the kernel parameters. Therefore, (Cristianini et al., 1999) proposed the Kernel-Adatron Algorithm, which can automatically perform model selection without being tested on a validation. Unfortunately, this algorithm is ineffective if the data have a flat ellipsoid distribution (Campbell, 2002). Therefore, one possible way is to consider the data distribution.

2.3. Genetic algorithms (GAs)

Evolutionary algorithms often have to solve optimization problems in the presence of a wide range of uncertainties (Yaochu & Branke, 2005). Genetic algorithms (GAs) are well suited for searching global optimal values in complex search space (multi-modal, multi-objective, non-linear, discontinuous, and highly constrained space), coupled with the fact that they work with raw objectives only when compared with conventional techniques (Holland, 1975; Goldberg, 1989; Waters & Sheble, 1993). For example, (Venkatraman & Yen, 2005) proposed a generic, two-phase framework for solving constrained optimization problems using GAs. Although many optimization methods have been proposed (e.g. Nelder-Mead simplex method), GAs are well suited to the concurrent manipulation of models with varying resolutions and structures since they can search non-linear solution spaces without requiring gradient information or a priori knowledge of model characteristics (Darwen & Xin, 1997; McCall & Petrovski, 1999). Based on fitness sharing, the learning system of GAs outperforms the tit-for-tat strategy against unseen test opponents. They learn using a "black box" simulation, with minimal prior knowledge of the learning task (Darwen & Xin, 1997).

In addition, the problem in binary coding lies in the fact that a long string always occupies the computer memory even though only a few bits are actually involved in the crossover and mutation operations. This is especially the case when a lot of parameters have to be adjusted in the same problem and a higher precision is required for the final result. This is also the main problem when initialing values of parameters of SVM in advance. To overcome this inefficient use of computer memory, the underlying real-valued crossover and mutation algorithm are employed (Huang & Huang, 1997). Contrary to the binary genetic algorithm (BGA), the real-valued genetic algorithm (RGA) uses real value as a