Applied Soft Computing 53 (2017) 308–322

Contents lists available at ScienceDirect

Applied Soft Computing

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

A modified support vector regression: Integrated selection of training subset and model

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

[a] School of Mathematics and Statistics, Xidian University, 266 Xinglong Section of Xifeng Road, Xi'an, Shaanxi 710126, China
[b] School of Science, NanChang Institute of Technology, NanChang 330099, JiangXi, China
[c] College of Mathematics and Information Science, Baoji University of Arts and Sciences, Baoji 721013, Shaanxi, China
[d] School of Public Health, Lanzhou University, Lanzhou 730000, China

ARTICLE INFO

Article history:
Received 22 October 2015
Received in revised form 20 June 2016
Accepted 29 December 2016
Available online 4 January 2017

Keywords:
Model selection
Optimal training subset
Nested particle swarm optimization
Support vector regression
Forecasting

ABSTRACT

In recent years, support vector regression (SVR) has become an emerging and popular forecasting technique in the field of machine learning. However, it is subjected to the model selection and learning complexity $O(K^*N^3)$, especially for a massive data set (N is the size of training dataset, and K is the number of search). How to simultaneously reduce K and N can give us insight and inspiration on designing an effective and accurate selection algorithm. To this end, this paper tries to integrate the selection of training subset and model for SVR, and proposes a nested particle swarm optimization (NPSO) by inheriting the model selection of the existing training subset based SVR (TS-SVR). This nested algorithm is achieved by adaptively and periodically estimating the search region of the optimal parameter setting for TS-SVR. Complex SVR, involving large-scale training data, can be seen as extensions of TS-SVRs, yielding a nested sequence of TS-SVRs with increasing sample size. The uniform design idea is transplanted to the above modeling process, and the convergence for the proposed model is proofed. By using two artificial regression problems, Boston housing and electric load in New South Wales as empirical data, the proposed approach is compared with the standard ones, the APSO-OTS-SVR, and other existing approaches. Empirical results show that the proposed approach not only can select proper training subset and parameter, but also has better generalization performance and fewer processing time.

1. Introduction

Support vector regression (SVR) is a very promising and popular forecasting model rooted in statistical learning theory [1], whose achievements are due to the linear or nonlinear kernel technique [2–4]. It is well-known that the generalization performance of SVR depends on the good choice of parameter setting [5,6]. For this purpose, the problem of SVR's parameter selection becomes a fundamental yet crucial task. It will be desirable to design an effective and accurate selection algorithm to make SVRs practical for the wide variety of practitioners in engineering applications of regression and modelling [7,4].

The parameter selection of SVR is known as the model selection problem, and can be formulated as an optimization problem of a function which is multimodal and only vaguely specified. A simple method to handle the model selection is to perform an exhaustive grid search over the parameter domain. Generally, the exhaustive grid search has high computational cost especially in large-scale training data set (the size of training dataset N is a large number): Firstly, the candidate size of the SVR parameter domain is large, so fine grid search is quite inefficient due to a large number of parameter combinations, becoming impracticable if the dimension of the parameter domain is large, even if the grid is not too fine [15]. Secondly, the computation complexity of the above problem is $O(K^*N^3)$ (N is the size of training dataset, and K is the number of search). To this end, the population based search algorithm, such as particle swarm optimization (PSO), is a good alternative for finding better tuning parameters [8,9]. Nevertheless, if the number of search K and the size of training dataset N are all considered, the complexity of the problem will still be significant for large-scale training data set. Based on the above rationality, how to simultaneously reduce K and N can give us insight and inspiration on designing an effective and accurate selection algorithm.

On the one hand, many researchers have studied how to reduce the number of trials in parameter combinations, i.e. K. By estimating some parameter-dependent error, four types of the gradient of parameter selection criterion are calculated to optimize the model

* Corresponding author.
E-mail addresses: ylyang@mail.xidian.edu.cn, youlongyang2015@163.com (Y. Yang).

http://dx.doi.org/10.1016/j.asoc.2016.12.053
1568-4946/

Nomenclature

K	the kernel function of SVR
δ or γ	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
d	the dimension of data set
$TS(m)$	a selected training subset with size m
$F_m(x)$, $F(x)$	the cumulative distribution function and empirical cumulative distribution function
$x_i(h)$	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
$P_{best}[TS(m)]$	the best particle of SVR for training subset $TS(m)$
n_k	the number of particles at the kth iteration
e_k	the minimum error at the kth iteration
$MaxN_k$	the maximum iteration number at the kth iteration

selection problem [10–13]. Although these algorithms can effectively reduce the K and show impressive gain in time complexity, it is likely that they get trapped in bad local minima. To avoid computing the derivatives of model selection criterion, Momma and Bennett introduce a pattern search algorithm which is suitable for optimizing SVR problems for which it is impossible or hard to obtain information about the derivatives [14]. Li et al. propose a multi-objective uniform design (MOUD) search algorithm to make the search scheme better uniformity and space-filling. This selection algorithm can dramatically reduce the K, avoid wasteful function evaluations of close-by patterns, then provide the flexibility to adjust the candidate set size under computational time constraint [15]. By combining the uniform design and stochastic optimization method, Jiménez et al. propose a random version of focused grid search [16], where more concentrated set in the parameter search space is repeatedly randomly screened and examined by using heuristic search. By updating the local and global best known positions, PSO can be expected to iteratively move the swarm toward the optimal solution, and perform a more effective search by combining the uniform design idea [17–20].

On the other hand, support vector regression (SVR) has high computation complexity $O(N^3)$ for large sample. Training data reduction is an effective method to reduce the N due to the sparseness of SVR. Based on it, researchers have proposed many effective solutions. One type of solution is to decompose the large dataset problem into several small sub-problems. By using the low-rank approximation method of dealing with the full kernel matrix, a reduced SVR (RSVR) is presented [21]. To further improve the computational efficiency, the other type of solution is to select a small-scale training subset for the original dataset. Lee et al. obtain a small-scale training subset by using the random sampling method, then train the SVR based on the training subset [22]. Brabanter et al. propose a modified active subset selection method by constructing a maximum objective function of quadratic Renyi entropy, and optimize the random training subsets using an iterative method, then determine an optimized fixed-size kernel models for large data sets [23]. The above SVRs for large scale of data are limited to the fixed-size training subset, Che constructs an approximation convexity optimization framework to determine the optimal size of training subset based SVR (TS-SVR), and solves it by a 0.618 method [24].

The work of literature [24] that hybridizes APSO and TS-SVR uses APSO for each model selection process of TS-SVRs. In particular, it is a two-step approach where the first stage is based on optimal training subset (OTS) algorithm to select a training subset considered in the second stage. For the TS-SVR with this training subset, an APSO algorithm is then employed to perform the model selection of the TS-SVR in the second stage. However, its model selection process is repeatedly performed for each iteration of training subset. This method does not consider the interconnection between training subset selection and model selection, which gives us the following motivation to achieve better performance: An integrated strategy between training subset and model selection is expected to make the selection process far more space-filling by cutting the wasteful parameter space of close-by patterns. As far as we know, no study truly integrates the training subset and model selection within a SVR modeling process.

To progressively exploit a wide candidate search region stage by stage, this paper tries to integrate the selection of training subset and model for SVR, and proposes a nested PSO by inheriting the model selection of the existing training subset based SVR. This modified SVR can simultaneously reduce both K and N by using the above integrated strategy, which is a fast SVR model suitable for large-scale training data. We firstly analyze and apply the training subset based SVR (TS-SVR) iteration process in literature [24] for exploring the parameter domain. Then, the movement region of the optimal parameter setting between two TSs with different sizes is estimated, which forms a nested mechanism and contracted search to dramatically reduce the candidate set space of parameter combinations. Finally, we connect the TS-SVRs from small to optimal size by a nested particle swarm optimization (NPSO) in order to search the tuning parameters periodically. Experimental results show that the proposed integrated SVR can select proper training subset and parameter with which the testing accuracy of trained SVRs is competitive to the standard ones, and the training time can be significantly shortened. The updated search regions can be adjusted adaptively and dynamically with the iteration process of TSs.

The rest of the study is organized as follows. Section 2 gives a brief introduction of the basic model. Section 3 presents the integrated training subset and model selection for SVR (I-TSMS-SVR). The convergence for the I-TSMS-SVR model is proofed in Section 4. Section 5 presents the numerical results. The final conclusion is drawn in Section 6.

2. The basic model

2.1. Support vector regression

In this subsection, a very brief description of ε-insensitive support vector regression (ε-SVR) is given, and a more thorough coverage is introduced in [1,25–27]. Specifically, the training data (x_i, y_i) of n records are given, where $(x_i, y_i) \in \mathbb{R}^d \times \mathbb{R}$. The SVR predictor will forecast record with input pattern $x \in \mathbb{R}^d$ by using a regression function of the form

$$f(x) = \sum_{i=1}^{n} \left(\alpha_i^* - \alpha_i\right) \mathbf{K}(x_i, x) + b \tag{1}$$

where $\mathbf{K} : \mathbb{R}^d \times \mathbb{R}^d \mapsto \mathbb{R}$ is the SVR kernel function that maps the input space into a higher-dimensional feature space, and this "kernel trick" makes the nonlinear regression on the input space be equivalent to the linear regression on the higher-dimensional feature space. Conventionally, the coefficients α_i^*, α_i are found by