ATLANTIS PRESS
Advances in Engineering Research (AER), volume 130
5th International Conference on Frontiers of Manufacturing Science and Measuring Technology (FMSMT 2017)

Optimization of Support Vector Regression Parameters by Flower Pollination Algorithm

Yuan Yang[1, a], Zhongqi Wang[1, b], Bo Yang[1, c] and Xudong Liu[1, d]

[1]The Ministry of Education Key Laboratory of Contemporary Design and Integrated Manufacturing Technology, Northwestern Polytechnical University, No. 127, Youyi Road (West), Xi'an 710072, China

[a]yangyuan0824@mail.nwpu.edu.cn, [b]wangzhqi@nwpu.edu.cn, [c]yohhanwen@mail.nwpu.edu.cn, [d]hurricane@mail.nwpu.edu.cn

Keywords: Support Vector Regression; Parameters Optimization; Flower Pollination Algorithm; GA; PSO.

Abstract. Support vector regression (SVR) is widely applied as a powerful method for data regression in engineering design and optimization. The regression accuracy and generalization performance of SVR model depend on the proper setting of its parameters. To this end, it is necessary to find an automated reliable, accurate and robust optimization approach to determining the optimal SVR parameter setting. This paper presents a SVR parameters optimization approach based on flower pollination algorithm (FPA), termed as FPA-SVR, to enhance the prediction ability of SVR model. Then a comparison is made among the performance of GA-SVR, PSO-SVR and FPA-SVR on one standard dataset. It can be concluded from the numerical results that the FPA-SVR model has superior regression accuracy and generalization performance.

Introduction

As a regression version of SVM [1], support vector regression (SVR) emerges as a powerful method to solve regression problems. However, there is an obvious problem appearing in the practical application of SVR. The problem is how to determinate the optimal parameters in SVR so that its performance can be enhanced. These parameters include the penalty parameter, loss function parameter, and the parameters in kernel function. To solve the problem mentioned above, many scholars and technicians have carried out a lot of research.

Traditionally, the methods of grid search or gradient descent algorithm are used to perform the optimization and usually based on cross validation to evaluate the performance of the SVM model. Recently, the evolutionary algorithms such as genetic algorithm (GA) and particle swarm optimization (PSO) are used for the optimization of SVM parameters since these algorithms are known to have better global search abilities. Wu et. al. [2] adopted a hybrid genetic algorithm (HGA) to search for the optimal type of kernel function and kernel parameters of SVR to increase its regression accuracy. Xin et. al. [3] employed GA to optimize the SVR parameters so as to improve the predicting ability of the models. In order to obtain an effective SVR model with optimal regression accuracy and generalization performance, Jiang et. al. [4] proposed GA, differential evolution algorithm (DEA) and PSO to determine optimal hyper-parameters of the SVR model. Hsieh et. al. [5] used PSO to search the optimal parameters for model selections in the hope of improving the performance of SVR. Aich and Banerjee [6] employed PSO for the purpose of optimizing SVM parameter combinations to develop the SVM model of electrical discharge machining process. Meng et. al. [7] applied PSO to optimize the hyper-parameters of SVR model and put forward a PSO-SVR model for forecasting coal seam gas content.

In this paper, in order to improve the regression accuracy and generalization performance of SVR models, an FPA based approach is proposed to optimizing the hyper-parameters in SVR.

Copyright © 2017, the Authors. Published by Atlantis Press.
This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

Support vector regression

Let x be mapped into a feature space by a nonlinear function $\varphi(x)$, the regression function can be expressed as Eq. 1.

$$f(x)=\omega\varphi(x)+b. \tag{1}$$

Here, ω and b are the coefficients of the function. Then SVR performs linear regression in the higher-dimensional feature space by ε-insensitive loss. Now, SVR can be formulated as Eq. 2.

$$\min_{\omega,b,\xi,\xi*} \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{n}\left(\xi_i+\xi_i^*\right)$$
$$\text{subject to}\begin{cases} y_i-\left(\omega\bullet\varphi(x_i)+b\right)\le\varepsilon+\xi_i \\ \left(\omega\bullet\varphi(x_i)+b\right)-y_i\le\varepsilon+\xi_i^* \\ \xi_i,\xi_i^*\ge 0, i=1,2,\cdots,n \end{cases}. \tag{2}$$

Here the penalty parameter $C>0$ determines the trade-off between the model complexity of $f(x)$ and the amount up to which deviations larger than ε are tolerated. ξ_i and ξ_i^* denote the slack variables that measure the error of the up and down sides, respectively. Then, by introducing Lagrange multipliers, the decision function given by Eq. 2 can be reformulated into Eq. 3.

$$f(x)=\sum_{i=1}^{n}\left(\alpha_i-\alpha_i^*\right)K(x_i,x)+b. \tag{3}$$

Here, α_i and α_i^* are the Lagrange multipliers and $K(x_i,x)$ represents the so called kernel function. In SVR, a suitable kernel function makes it possible to map a nonlinear input space to a higher-dimensional feature space where linear regression can be performed [1]. Generally, using Gaussian function will yield better prediction performance. In this paper, our focus is also put on the widely used kernel function of radial basis function (RBF), which is defined as Eq. 4.

$$K(x_i,x)=\exp\left(-\gamma\|x_i-x\|^2\right). \tag{4}$$

Here γ is the width of the Gaussian function.

The determination of the optimal hyper-parameters is an important step in SVR construction. The parameters include the penalty parameter *C*, the RBF kernel function width γ, and the tube size of ε-insensitive loss function. In this paper, an FPA based optimization approach is proposed to determining the hyper-parameters (*C*, γ and ε) of SVR, termed as FPA-SVR, which simultaneously optimizes all SVR parameters from the training data.

The goal of optimizing parameters for SVR is to utilize optimized procedures to find the parameters that minimize the predicting error, defined as RRMSE (relative root mean squared error), of the testing data. Supposing there is a testing set $S=\{(\mathrm{x}_i,y_i)|i=1,\cdots,N_{test}\}$, where x_i is the *i*-th input data and y_i is the corresponding output data. N_{test} is number of the input data in the testing set. $\hat{y}_i$ represents the predicting output of the SVR model for the *i*-th input data in the testing set. According to the proposed FPA-SVR model, the evaluation function, RRMSE, is chosen to optimize the parameters by using FPA. Thus the final optimization model can be presented as Eq. 5.

$$\begin{cases} \text{Find: } P=\left[C,\varepsilon,\gamma\right] \\ \text{Min: RRMSE}=\dfrac{\sqrt{\dfrac{1}{N_{test}}\sum_{i=1}^{N_{test}}\left(y_i-\hat{y}_i\right)^2}}{\dfrac{1}{N_{test}}\sum_{i=1}^{N_{test}}\left(y_i\right)}. \\ \text{s.t. } C>0, \varepsilon>0, \gamma>0 \end{cases} \tag{5}$$

Flower pollination algorithm

Inspired by the flower pollination characteristics, the flower pollination algorithm was proposed by Yang [8]. For simplicity, the following five rules are summarized.

Rule 1: Biotic and cross-pollination can be regarded as global pollination process, and pollinators transfer pollen via Lévy flights.

Rule 2: Abiotic and self-pollination can be regarded as local pollination process.

Rule 3: Flower constancy can be regarded as a reproduction probability that is proportionate to the similarity of two flowers involved.

Rule 4: A switch probability $p\in\left[0,1\right]$ is used to control the shift between local and global pollination.

Rule 5: Each plant has only one flower, and each flower produce only one pollen grain.

According to the rules above, the FPA can be represented mathematically as follows:

In the global pollination step:

$$\mathbf{x}_i^{t+1}=\mathbf{x}_i^t+\gamma L\left(\lambda\right)\left(\mathbf{g}_*-\mathbf{x}_i^t\right). \tag{6}$$

Here $\mathbf{x}_i^t$ is the pollen i or solution vector $\mathbf{x}_i$ at generation t. $\mathbf{g}_*$ is the current best solution found among all solutions at the current generation. γ is a scaling factor to control the step size. $L\left(\lambda\right)$ is the strength of the pollination, which essentially is a Lévy flight based step size.

$$L\left(\lambda\right)\sim\frac{\lambda\Gamma\left(\lambda\right)\sin\left(\pi\lambda/2\right)}{\pi}\frac{1}{s^{1+\lambda}},\left(s\gg s_0>0\right). \tag{7}$$

Here, $\Gamma\left(\lambda\right)$ is the gamma function and parameters $\lambda=1.5$ and $s_0=0.1$ are as suggested by Yang [8].

In the local pollination step:

$$\mathbf{x}_i^{t+1}=\mathbf{x}_i^t+\epsilon\left(\mathbf{x}_j^t-\mathbf{x}_k^t\right). \tag{8}$$

Here, $\mathbf{x}_j^t$ and $\mathbf{x}_k^t$ are the pollen grains form different flowers of the same plant species. ϵ is a local random walk drawn from a uniform distribution in $\left[0,1\right]$.

FPA-SVR model

In order to perform this optimization, the FPA is used to find a better combination of the three parameters in the SVR so that a smaller RRMSE is obtained in each predicting generation. Fig. 1

illustrates the flowchart of the proposed FPA-SVR model. As described above, the framework of the FPA-SVR model can be summarized into three stages.

In the first stage, the original input data is preprocessed by scaling the training data set and testing data set. Generally, all the values of the data sets should be linearly scaled into the range of $[0,1]$.

In the second stage, the hyper-parameters (C, γ and ε) of SVR are selected carefully with the FPA. By FPA operations, new solutions are generated by global pollination and local pollination while the proposed solution (the hyper-parameters C, γ and ε) is accepted or not depends on the fitness (RRMSE) of the solution. This process is repeated until the maximum generation number is reached.

Finally, according to the effective FPA-SVR model, an experiment can be carried out to verify the proposed approach by using standard regression dataset. The data set is obtained from University of California at Irvin (UCI) Machine Learning Repository [9]. The experiment is carried out and implemented on the Matlab 8.1.0.604 (2013a) development environment by extending the LIBSVM toolbox which is originally designed by Chang and Lin [10].

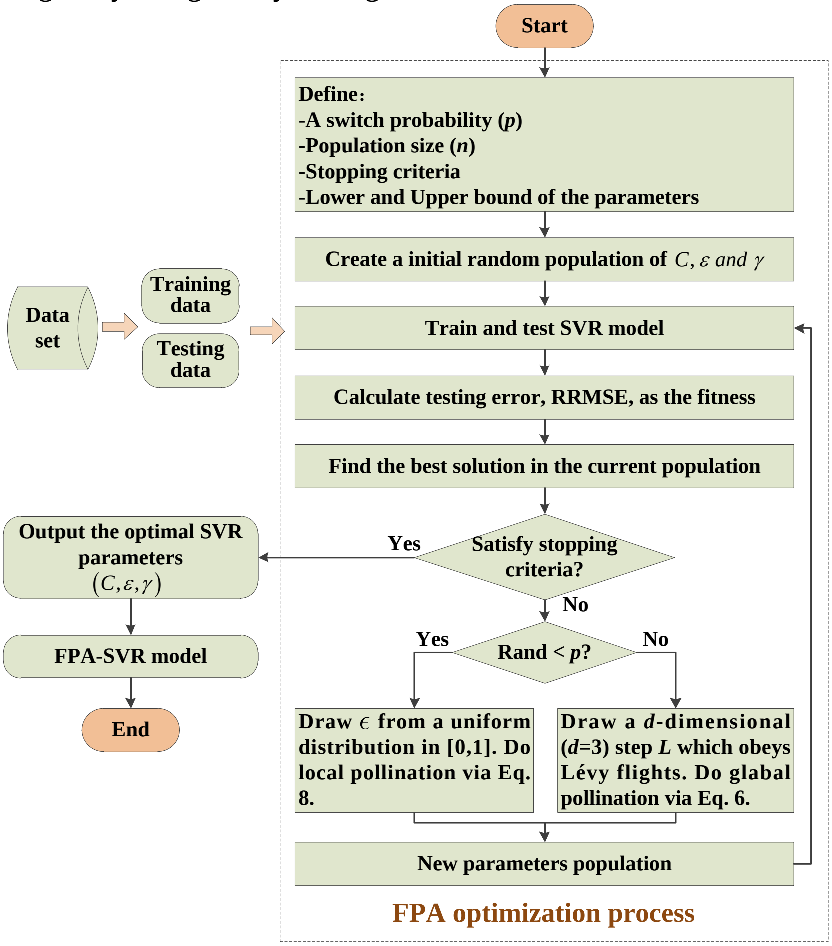

Fig. 1 Flowchart of the FPA based SVR parameter optimization procedure

Experiment

The dataset, extracted from UCI database and created by Angeliki [11], comprises 768 samples and 8 features, aiming to predict a real valued response. In this experiment, 618 samples are chosen as training set randomly, while the remaining 150 samples are used as testing set. After tens of trials, the population size is fixed as 20 and the maximum generation number is set as 100. Then GA, PSO and PFA algorithms are proposed to seek the corresponding optimal hyper-parameters (C, γ and ε) of SVR models. The initial randomly generated population is set the same for all the three evolutionary algorithms. The selected optimum setting values of the other parameters of the algorithms above are given as follows:

GA setting: Crossover probability $P_c = 0.3$ and mutation probability $P_m = 0.02$.

PSO setting: Cognitive and social scaling parameters $c_1 = c_2 = 2$.

FPA setting: The switch probability $p = 0.6$.

Table 1 lists the optimization results and the suitable parameters for the different SVR models. The performance of GA, PSO and FPA is shown in Fig. 2.

Table 1 Optimized parameters and the corresponding RRMSE

Optimization model	Parameters			RRMSE
	C	ε	γ	
GA-SVR	6.9665×10^7	0.0207	0.1702	3.00%
PSO-SVR	5.4734×10^7	0.0135	0.4191	2.86%
FPA-SVR	8.2718×10^7	0.0136	0.3183	2.74%

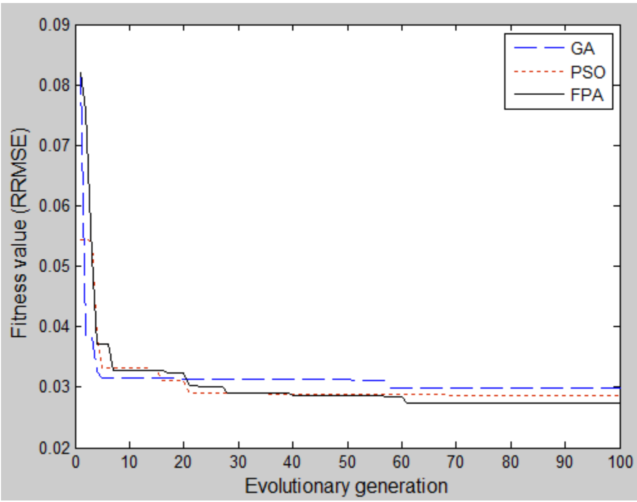

Fig. 2 The minimum fitness values by GA, PSO and FPA

It can be seen that, among the three parameters optimization methods, the FPA-SVR performs better than GA-SVR and PSO-SVR methods because its predicted solutions are much closer to the expected distributions. The test results (See Table 1) showed that the FPA-SVR model yields improved prediction results and significantly outperforms the other two prediction models.

Conclusion and future research

SVR is widely used in different applications. However, the values of SVR parameters affect the regression accuracy and generalization performance of SVR models. In this paper, a flower pollination algorithm based approach (FPA-SVR) is proposed to search the SVR parameters that minimize the testing error RRMSE. The main conclusion of the work can be summarized as: An FPA-SVR model is constructed by integrating FPA and SVR. FPA can be used to find the best hyper-parameters of SVR, and the proposed FPA-SVR is a relatively effective model in terms of accuracy and efficiency in comparison with GA-SVR and PSO-SVR models.

Three directions for future study are suggested. First, an RBF kernel function is pre-assumed in the SVR model in this paper. However, the type of kernel function and the other parameter values of different kernel functions can also be optimized using the same approach. Second, the experiment in this paper is performed using only one dataset, but more other public datasets or real-world problems should be tested in the future to verify and extend the proposed FPA-SVR model. Third, due to the performance of FPA, it would be worth exploring the potential application of FPA to the support vector classification.

Acknowledgements

This work is supported by National Natural Science Foundation of China (Grant No. 51375396) and Natural Science Foundation of Shannxi Province, China (Grant No. 2016JM5011).

References

[1] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer, New York, USA, 1995.

[2] C. H. Wu, G. H. Tzeng, R. H. Lin, A novel hybrid genetic algorithm for kernel function and parameter optimization in support vector regression, *Expert Systems with Applications*, vol. 36, no. 3, pp. 4725-4735, 2009.

[3] N. Xin, X. Gu, H. Wu, Y. Hu, Z. Yang, Application of genetic algorithm-support vector regression (GA-SVR) for quantitative analysis of herbal medicines, *Journal of Chemometrics*, vol. 26, no. 7, pp. 353–360, 2012.

[4] M. Jiang, L. Zhu, Y. Wang, W. Huang, H. Zhang, Study on parameter optimization for support vector regression in solving the inverse ECG problem, *Computational and Mathematical Methods in Medicine*, vol. 2013, no. 4, pp. 529-543, 2013.

[5] H. Hsieh, T. P. Lee, T. S. Lee, A hybrid particle swarm optimization and support vector regression model for financial time series forecasting, *International Journal of Business Administration*, vol. 2, no. 2, pp. 48-56, 2011.

[6] U. Aich and S. Banerjee, Modeling of EDM responses by support vector machine regression with parameters selected by particle swarm optimization, *Applied Mathematical Modelling*, vol. 38, no. 11-12, pp. 2800-2818, 2014.

[7] Q. Meng, X. Ma, Y. Zhou, Forecasting of coal seam gas content by using support vector regression based on particle swarm optimization, *Journal of Natural Gas Science and Engineering*, vol. 21, no. 21, pp. 71-78, 2014.

[8] X. S. Yang, Flower pollination algorithm for global optimization, *Unconventional Computation and Natural Computation*, Springer Berlin Heidelberg, vol. 7445, pp. 240 - 249, 2012.

[9] M. Lichman, UCI Machine Learning Repository [http://archive.ics.uci.edu/ml], Irvine, CA: University of California, School of Information and Computer Science, 2013.

[10] C. C. Chang and C. J. Lin, LIBSVM: A library for support vector machines, *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 1-27, 2011.

[11] A. Tsanas and A. Xifara, Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools, *Energy and Buildings*, vol. 49, pp. 560-567, 2012.