Hyper parameters Optimization of Support Vector Regression based on a Chaotic Pigeon-Inspired Optimization Algorithm

Niam Abdulmunim Al-Thanoon
Department of Operations Research and Intelligent Techniques, University of Mosul, Mosul, Iraq
E-mail: niam.munim@uomosul.edu.iq

Zakariya Yahya Algamal
Department of Statistics and Informatics, University of Mosul, Mosul, Iraq
E-mail: zakariya.algamal@uomosul.edu.iq
ORCID: 0000-0002-0229-7958

Omar Saber Qasim
Department of Mathematics, University of Mosul, Mosul, Iraq
E-mail: omar.saber@uomosul.edu.iq
ORCID: 0000-0003-3301-6271

Article Info
Page Number: 4997 - 5008
Publication Issue:
Vol 71 No. 4 (2022)

Article History
Article Received*: 25 March 2022*
Revised*: 30 April 2022*
Accepted*: 15 June 2022*
Publication*: 19 August 2022*

Abstract

Support vector regression (SVR) as a data mining tool has been applied in several real problems. However, it is usually needed to tune manually the hyperparameter. Meta-heuristic algorithms have been used as hyperparameters tuning procedure. In this paper, chaotic pigeon-inspired optimization algorithm is proposed to enhance the exploration and exploitation capability of the pigeon-inspired optimization algorithm.The results, which were applied to four of the datasets, show that the proposed algorithm has the possibility to obtain results better than the cross-validation method through prediction values and time spent, as well as the efficiency of the proposed algorithm in improving prediction and computational time spent in processing, compared to algorithms inspired by nature.

Keywords: pigeon-inspired optimization algorithm, evolutionary algorithm, SVR.

1. Introduction

Support vector machine (SVM) is one of the most important techniques that have received great interest from researchers because of its practical advantages that have improved its performance in many applications of classification and regression. (Chuang and Lee 2011). Originally, “SVM technique was used to solve problems related to classification. With the introduction of this technique's insensitive loss function by Vapnik, the concept of SVM technology has been extended to include solutions related to nonlinear regression, called support vector regression (SVR)(Chuang and Lee 2011; Ye et al. 2017; Zhao and Sun 2010).

SVR has three advantages: (1) guaranteed convergence to optimal solutions because quadratic programming is used with linear constraints for learning the data. (2) Computationally efficient for nonlinear relationship modeling by using kernel mapping. And, (3) a superior generalization performance (lower error rates on test set) (Xu et al. 2018).

The accuracy of the results for SVR is highly dependent on a number of hyperparameters which greatly influence the finding of optimal solutions. In search space exhaustive search is usually used to investigate all hyperparameter combinations in order to increase the predictive power of SVR (Kaneko and Funatsu 2015). Despite the characteristics of SVR technique, there are still a number of problems related to choosing the optimal model, including the feature selection process. In other words, SVR technology cannot perform the feature selection process (Al-Thanoon et al. 2018).

The nature-inspired algorithms, have attracted a large number of people interested in classification and prediction, and have achieved good and competitive results when finding solutions related to optimization, including the problem of obtaining the best values for hyperparameters.(Chou and Pham 2017; Laref et al. 2019; Li et al. 2018)In the literature, there are many studies on the tuning of hyperparameters of SVR through nature-inspired algorithms, such as (Cheng et al. 2007; Cheng et al. 2009; Hong et al. 2011; Huang 2012; Kazem et al. 2013; Li et al. 2018; Nait Amar and Zeraibi 2018; Üstün et al. 2005; Wu et al. 2009; Zhang et al. 2017). In recent years, researchers have developed a several number of algorithms inspired by nature to improve and enhance processes related to exploration and exploitation. Among these important algorithms is a pigeon-inspired optimization algorithm that is highly efficient”.

The main objective of this paper is to optimize the hyperparameters of SVR teqnique by improving the pigeon-inspired optimization algorithm. The efficiency of the proposed algorithm in this study was measured with a number of other previous algorithms.

2. Support vector regression (SVR)

SVM technique has been used to solve many different classification problems. However, with the introduction of the insensitive loss function for by Vapnik (1999), “the work of the SVM technique was extended to solve problems related to nonlinear regression, which is called the support vector regression technique. In SVR, the target variable is a quantitative variable, such as chemical activity (Fu et al. 2011), spectral analysis (Malik et al. 2014; Xu et al. 2018), and stock price forecasting.

Given a training dataset of n observations $\{(\mathbf{x}_i, y_i)\}_{i=1}^{n}$, where $\mathbf{x}_i = (x_{i,1}, x_{i,2}, ..., x_{i,p}) \in \mathrm{R}^p$ represents a vector of the i^{th} feature, $y_i \in \mathrm{R}$ for $i = 1,...,n$ is The target variable, which is one of the quantitative variables, and an insensitive loss function, the SVR technique can be obtained by solving the following optimization problems

$$\min_{\mathbf{w},b}\left\{\frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{n}\left(\zeta_i + \tilde{\zeta}_i\right)\right\}$$
$$\text{S.T.}\quad \begin{cases} y_i - (\mathbf{w}\cdot\varphi(\mathbf{x}_i) + b) \le \varepsilon + \tilde{\zeta}_i \\ (\mathbf{w}\cdot\varphi(\mathbf{x}_i) + b) - y_i \le \varepsilon + \zeta_i \\ \zeta_i, \tilde{\zeta}_i \ge 0, \end{cases} \tag{1}$$

where $C > 0$ represents the penalized parameter, ζ_i and $\tilde{\zeta}_i$ are slack variables, $\varphi(\mathbf{x}_i)$ representsthe nonlinear mapping, $\mathbf{w}$ is the weight vector and b is the bias.

Then, Eq. (1) solved through Lagrangian multipliers as

$$\min_{\tilde{\alpha},\alpha}\frac{1}{2}\sum_{i,j=1}^{n}(\tilde{\alpha}_i - \alpha_i)(\tilde{\alpha}_j - \alpha_j)K(\mathbf{x}_i, \mathbf{x}_j) + \varepsilon\sum_{i=1}^{n}(\tilde{\alpha}_i - \alpha_i) - \sum_{i=1}^{n} y_i(\tilde{\alpha}_i - \alpha_i)$$
$$\text{S.T.}\quad \begin{cases} \sum_{i=1}^{n}(\alpha_i - \tilde{\alpha}_i) = 0 \\ 0 \le \alpha_i, \tilde{\alpha}_i \le C, \end{cases} \tag{2}$$

where $K(\mathbf{x}_i, \mathbf{x}_j)$ stands for kernel mapping, and α_i, $\tilde{\alpha}_i$ are Lagrangian multipliers. The regression hyperplane for the underlying regression problem is then given by

$$y_i = f(\mathbf{x}_i) = \sum_{\mathbf{x}_i = \mathrm{SV}} (\tilde{\alpha}_i + \alpha_i) K(\mathbf{x}_i, \mathbf{x}_j) + b, \quad (3)$$

where SV represents the support vectors.

3. Pigeon-inspired optimization

The pigeon-inspired algorithm (POA) mainly consists of two operators: the map and compass operator and the landmark operator(Cheng et al. 2019; Duan and Qiu 2019; Jiang et al. 2019; Qiu and Duan 2020; Xiang et al. 2019; Yang et al. 2019; Zhong et al. 2019). Pigeons sense the geomagnetic field in the map trigger and compass to form the maps for homing. Suppose the required search space consists of N dimensions, then the first bath can be represented by an N-dimensional vector $X_i = (X_{i,1}, X_{i,2}, ..., X_{i,\mathrm{N}})$. The velocity in the pigeon algorithm, by changing the position of this bath, can be represented by a second vector consisting of N-dimensional $V_i = (V_{i,1}, V_{i,2}, ..., V_{i,\mathrm{N}})$. The best previously visited position of the i-th pigeon is denoted as $P_i = (P_{i,1}, P_{i,2}, ..., P_{i,\mathrm{N}})$.The global best position of the swarm is $g = (g_1, g_2, \ldots, g_N)$. The flying in pigeon algorithmcan be represents as:

$$V_i(t+1) = V_i(t) \times e^{-Rt} + \tau \times (X_g - X_i(t)) \quad (4)$$

$$X_i(t+1) = X_i(t) + V_i(t+1), \quad (5)$$

where R is a map and compass factor, while τ is a uniform random number in the range $[0, 1]$, X_g denotes the global best solution, $X_i(t)$ is the current position of a pigeon at instance t, and $V_i(t)$ is the current velocity at iteration t.

In landmark operator, pigeons are categorized by their fitness. In each new generation, a number of pigeon positions are updated by Eq. (4), where half the number of pigeons is counted on to calculate the desired position of the centered pigeon, while all the other pigeons change and adjust their destination

$$N_p(t+1) = \frac{N_p(t)}{2}, \quad (6)$$

where N_p denoted number of pigeons in iteration t.

The positions for the desired destination are calculated by Eq. (5), while every other pigeon changes and modifies its position towards this position through Eq. (6) (Duan and Qiao 2014).

$$X_c(t+1) = \frac{\sum X_i(t+1) \times \text{Fitness}(X_i(t+1))}{N_p \sum \text{Fitness}(X_i(t+1))} \quad (7)$$

$$X_i(t+1) = X_i(t) + rand \times \left(X_c(t+1) \times X_i(t)\right), \quad (8)$$

where X_c is the position of the centered pigeon (desired destination)".

4. Proposed method

In SVR there are many important parameters that are adjustable for optimizing solutions. These parameters are called hyperparameters, such as,ε-insensitive loss function, ε ,penalized parameter C , and the kernel parameter. "The performance of SVR is described as being very sensitive to the selection of hyperparameters and there are no special mathematical procedures to derive the desired values well(Tsirikoglou et al. 2017).

In the literature, there are many serious attempts to improve the performance of SVR by selecting the best possible values for hyperparameters (Cherkassky and Ma 2004; Chou and Pham 2017; Ito and Nakano 2003; Laref et al. 2019; Li et al. 2018; Tsirikoglou et al. 2017). Nature-inspired algorithms are among the most important algorithms used for selecting hyperparameters of SVR(Cao and Wu 2016; Cheng et al. 2007; Cheng et al. 2009; Hong et al. 2011; Huang 2012; Kazem et al. 2013; Li et al. 2018; Nait Amar and Zeraibi 2018; Üstün et al. 2005; Wu et al. 2009; Zhang et al. 2017). Our contribution to this study is by optimizing the hyperparameters of the SVR technology using the pigeon-inspired algorithm. In this study we suggested the type of kernel function is Gaussian kernel with parameter $\sigma > 0$.

Chaos theory describes non-uniform mathematical behavior in many nonlinear systems and for this reason, chaotic maps are used. Particles can travel as chaotic maps in a limited range of systems(Sayed et al. 2018). Anarchy maps are implemented to improve the quality of searching for optimal solutions. Considering that the feature selection problem is one of the important problems to improve classification efficiency with search range [0, 1], chaotic maps can be used to improve this problem. (Sayed et al. 2018).

In this paper, chaotic maps are proposed to improve the performance of the pigeon-inspired algorithm in terms of avoiding being trapped for determining the hyperparameters of SVR by embedding in τ. Ten chaotic maps are used in this paper. The description of these maps is explained in Table 1.

Table 1: The description of the most used maps

Name	Definition	Range
Ch1=Chebyshev	$x_{k+1} = \cos(k \cos^{-1}(x_k))$	(-1,1)
Ch2=Circle	$x_{k+1} = \text{mod}(x_k + 0.2 - \frac{0.5}{2\pi}\sin(2\pi x_k),1)$	(0,1)
Ch3=Guass/mouse	$x_{k+1} = \begin{cases} 1 & \mathrm{x}_k = 0 \\ \frac{1}{\text{mod}(x_k,1)}, otherwise \end{cases}$	(0,1)
Ch4=Iterative	$x_{k+1} = \sin\left(\frac{(0.7)\pi}{x_k}\right)$	(-1,1)
Ch5=Logistic	$x_{k+1} = 4x_k(1-x_k)$	(0,1)
Ch6=Piecewise	$x_{k+1} = \begin{cases} \frac{x_k}{0.4} & 0 \le x_k < 0.4 \\ \frac{x_k - 0.4}{0.1} & 0.4 \le x_k < 0.5 \\ \frac{0.6 - x_k}{0.1} & 0.5 \le x_k < 0.6 \\ \frac{1 - x_k}{0.4} & 0.6 \le x_k < 1 \end{cases}$	(0,1)
Ch7=Sine	$x_{k+1} = \sin(\pi x_k)$	(0,1)
Ch8=Singer	$x_{k+1} = 1.07(7.86x_k - 23.31(x_k)^2 + 28.75(x_k)^3 - 13.$	(0,1)
Ch9=Sinusoidal	$x_{k+1} = 2.3x_k \sin(\pi x_k)$	(0,1)
Ch10=Tent	$x_{k+1} = \begin{cases} \frac{x_k}{0.7} & x_k < 0.7 \\ \frac{10}{3}(1-x_k) & x_k \ge 0.7 \end{cases}$	(0,1)

The steps of our proposed method are:

Step 1: The number of pigeons is $np = 30$ and the max No. of iterations is $t_{max} = 100$.

Step 2: The first three positions represent the hypermeters are randomly generated as $C \sim U(0, 6.5)$, $\sigma \sim U(0,3)$, and $\varepsilon \sim U(0,1)$.

Step 3: The fitness function is defined as

$$\text{fitness} = \min\left[\frac{1}{n_{test}}\sum_{i=1}^{n_{test}}(y_{i,test} - \hat{y}_{i,test})^2\right], \quad (9)$$

where the fitness is calculated for the testing dataset.

Step 4: Steps 2 and 3 are repeated until a $t_{\max}$ is reached".

5. Results

To test the predictive performance of the proposed method, a number of experiments are used to compare with standard POA and cross-validation with tenfold (CV) approach. Various chemical data sets were used in this research: "Antidiabetic activity of dipeptidyl peptidase-IV inhibitors (Dataset 1) (Al-Fakih et al. 2019) and influenza neuraminidase a/PR/8/34 (H1N1) inhibitors (Dataset 2).Datasets include a large number of descriptors as features. Table 2 provides an illustrative summary of the data set used. Each data set used in this study was divided into two sets: the training data set consisting of 70% of the samples, and the test data set consisting of 30%. This division is repeated in each step. Two important evaluation criteria were used: the mean squared error (MSE) of the training dataset ($MSE_{train} = \sum_{i=1}^{n_{train}}(y_{i,train} - \hat{y}_{i,train})^2 / n_{train}$) and MSE of the testing dataset ($MSE_{test} = \sum_{i=1}^{n_{test}}(y_{i,test} - \hat{y}_{i,test})^2 / n_{test}$)".

Table 2: The datasets used.

Dataset type	#Sample	#features
D1	134	1048
D2	479	2881

Table 3: The prediction performance of the used algorithms

Algorithms	Dataset 1		Dataset 2	
	MSE_{train}	MSE_{test}	MSE_{train}	MSE_{test}
Ch1	7.483	7.84	8.791	9.148
Ch2	8.391	8.748	9.699	10.056
Ch3	9.208	9.565	10.516	10.873
Ch4	7.367	7.724	8.675	9.032
Ch5	**6.084**	**6.441**	**7.392**	**7.749**
Ch6	7.952	8.309	9.26	9.617
Ch7	8.887	9.244	10.195	10.552
Ch8	8.637	8.994	9.945	10.302
Ch9	9.147	9.504	10.455	10.812
Ch10	9.627	9.984	10.935	11.292
POA	10.771	11.128	12.079	12.436
CV	12.541	12.898	13.849	14.206

Table 4: The averaged computational time.

Algorithms	Dataset 1	Dataset 2
Ch1	143.36	147.44
Ch2	145.67	149.75
Ch3	144.44	148.52
Ch4	157.13	161.21
Ch5	**139.47**	**143.55**
Ch6	153.2	157.28
Ch7	154.44	158.52
Ch8	153.96	158.04
Ch9	156.67	150.75
Ch10	148.53	152.61
POA	166.21	170.29

CV	190.587	201.338

6. Conclusion

To improve the exploration and exploitation of POA used to get the optimum hyperparameters of the SVR chaotic POA algorithm, several chaotic maps are used in this paper to improve the working efficiency of POA. Experimental numerical results on two sets of chemical data showed that the proposed chaotic performance of POA compared with POA and standard CV leads to better prediction performance and less computational time. In addition, the performance of the Ch5 chaotic map was superior to the rest of the maps used.

7. References

1. Al-Fakih A, Algamal Z, Lee M, Aziz M, Ali H (2019) A QSAR model for predicting antidiabetic activity of dipeptidyl peptidase-IV inhibitors by enhanced binary gravitational search algorithm SAR and QSAR in Environmental Research 30:403-416
2. Al-Thanoon NA, Qasim OS, Algamal ZY (2018) Tuning parameter estimation in SCAD-support vector machine using firefly algorithm with application in gene selection and cancer classification Comput Biol Med 103:262-268 doi:10.1016/j.compbiomed.2018.10.034
3. Cao G, Wu L (2016) Support vector regression with fruit fly optimization algorithm for seasonal electricity consumption forecasting Energy 115:734-745
4. Cheng C-T, Wang W-C, Xu D-M, Chau KW (2007) Optimizing Hydropower Reservoir Operation Using Hybrid Genetic Algorithm and Chaos Water Resources Management 22:895-909 doi:10.1007/s11269-007-9200-1
5. Cheng J, Qian J, Guo Y-n Adaptive chaotic cultural algorithm for hyperparameters selection of support vector regression. In: International Conference on Intelligent Computing, 2009. Springer, pp 286-293
6. Cheng S, Lei X, Lu H, Zhang Y, Shi Y (2019) Generalized pigeon-inspired optimization algorithms Science China Information Sciences 62 doi:10.1007/s11432-018-9727-y
7. Cherkassky V, Ma Y (2004) Practical selection of SVM parameters and noise estimation for SVM regression Neural networks 17:113-126

8. Chou J-S, Pham A-D (2017) Nature-inspired metaheuristic optimization in least squares support vector regression for obtaining bridge scour information Information Sciences 399:64-80 doi:10.1016/j.ins.2017.02.051
9. Chuang C-C, Lee Z-J (2011) Hybrid robust support vector machines for regression with outliers Applied Soft Computing 11:64-72
10. Duan H, Qiao P (2014) Pigeon-inspired optimization: a new swarm intelligence optimizer for air robot path planning International Journal of Intelligent Computing and Cybernetics 7:24-37 doi:10.1108/ijicc-02-2014-0005
11. Duan H, Qiu H (2019) Advancements in pigeon-inspired optimization and its variants Science China Information Sciences 62 doi:10.1007/s11432-018-9752-9
12. Fu GH, Cao DS, Xu QS, Li HD, Liang YZ (2011) Combination of kernel PCA and linear support vector machine for modeling a nonlinear relationship between bioactivity and molecular descriptors Journal of chemometrics 25:92-99
13. Hong W-C, Dong Y, Chen L-Y, Wei S-Y (2011) SVR with hybrid chaotic genetic algorithms for tourism demand forecasting Applied Soft Computing 11:1881-1890 doi:10.1016/j.asoc.2010.06.003
14. Huang C-F (2012) A hybrid stock selection model using genetic algorithms and support vector regression Applied Soft Computing 12:807-818 doi:10.1016/j.asoc.2011.10.009
15. Ito K, Nakano R Optimizing support vector regression hyperparameters based on cross-validation. In: Proceedings of the International Joint Conference on Neural Networks, 2003., 2003. IEEE, pp 2077-2082
16. Jiang F, He J, Tian T (2019) A clustering-based ensemble approach with improved pigeon-inspired optimization and extreme learning machine for air quality prediction Applied Soft Computing 85:105827 doi:10.1016/j.asoc.2019.105827
17. Kaneko H, Funatsu K (2015) Fast optimization of hyperparameters for support vector regression models with highly predictive ability Chemometrics and Intelligent Laboratory Systems 142:64-69 doi:10.1016/j.chemolab.2015.01.001
18. Kazem A, Sharifi E, Hussain FK, Saberi M, Hussain OK (2013) Support vector regression with chaos-based firefly algorithm for stock market price forecasting Applied Soft Computing 13:947-958 doi:10.1016/j.asoc.2012.09.024

19. Laref R, Losson E, Sava A, Siadat M (2019) On the optimization of the support vector machine regression hyperparameters setting for gas sensors array applications Chemometrics and Intelligent Laboratory Systems 184:22-27 doi:10.1016/j.chemolab.2018.11.011
20. Li S, Fang H, Liu X (2018) Parameter optimization of support vector regression based on sine cosine algorithm Expert Systems with Applications 91:63-77 doi:10.1016/j.eswa.2017.08.038
21. Malik B, Chaitanya K, Benaissa M (2014) Support vector regression with digital band pass filtering for the quantitative analysis of near-infrared spectra Journal of Chemometrics 28:116-122
22. Nait Amar M, Zeraibi N (2018) Application of hybrid support vector regression artificial bee colony for prediction of MMP in CO2-EOR process Petroleum doi:10.1016/j.petlm.2018.08.001
23. Qiu H, Duan H (2020) A multi-objective pigeon-inspired optimization approach to UAV distributed flocking among obstacles Information Sciences 509:515-529 doi:10.1016/j.ins.2018.06.061
24. Sayed GI, Darwish A, Hassanien AE (2018) A New Chaotic Whale Optimization Algorithm for Features Selection Journal of Classification 35:300-344 doi:10.1007/s00357-018-9261-2
25. Tsirikoglou P, Abraham S, Contino F, Lacor C, Ghorbaniasl G (2017) A hyperparameters selection technique for support vector regression models Applied Soft Computing 61:139-148
26. Üstün B, Melssen W, Oudenhuijzen M, Buydens L (2005) Determination of optimal support vector regression parameters by genetic algorithms and simplex optimization Analytica Chimica Acta 544:292-305
27. Vapnik VN (1999) An overview of statistical learning theory IEEE transactions on neural networks 10:988-999
28. Wu C-H, Tzeng G-H, Lin R-H (2009) A Novel hybrid genetic algorithm for kernel function and parameter optimization in support vector regression Expert Systems with Applications 36:4725-4735 doi:10.1016/j.eswa.2008.06.046

29. Xiang S, Xing L, Wang L, Zou K (2019) Comprehensive learning pigeon-inspired optimization with tabu list Science China Information Sciences 62 doi:10.1007/s11432-018-9728-x
30. Xu S, Lu B, Baldea M, Edgar TF, Nixon M (2018) An improved variable selection method for support vector regression in NIR spectral modeling Journal of Process Control 67:83-93
31. Yang Z, Liu K, Fan J, Guo Y, Niu Q, Zhang J (2019) A novel binary/real-valued pigeon-inspired optimization for economic/environment unit commitment with renewables and plug-in vehicles Science China Information Sciences 62 doi:10.1007/s11432-018-9730-4
32. Ye Y-F, Shao Y-H, Deng N-Y, Li C-N, Hua X-Y (2017) Robust Lp -norm least squares support vector regression with feature selection Applied Mathematics and Computation 305:32-52 doi:10.1016/j.amc.2017.01.062
33. Zhang J, Zheng C-H, Xia Y, Wang B, Chen P (2017) Optimization enhanced genetic algorithm-support vector regression for the prediction of compound retention indices in gas chromatography Neurocomputing 240:183-190 doi:10.1016/j.neucom.2016.11.070
34. Zhao Y-P, Sun J-G (2010) Robust truncated support vector regression Expert Systems with Applications 37:5126-5133
35. Zhong Y, Wang L, Lin M, Zhang H (2019) Discrete pigeon-inspired optimization algorithm with Metropolis acceptance criterion for large-scale traveling salesman problem Swarm and Evolutionary Computation 48:134-144 doi:10.1016/j.swevo.2019.04.002