Academic Journal of Computing & Information Science
ISSN 2616-5775 Vol. 4, Issue 5: 1-7, DOI: 10.25236/AJCIS.2021.040501

Prediction of global electricity generation by using least squares support vector regression with sparrow search algorithm

Yuanjie Xu

North University of China, School of Electrical and Control Engineering, Taiyuan, Shanxi, 030051, China

Abstract: *In order to accurately predict the development of global electricity generation, this paper presents a prediction model of least squares support vector machine (LS-SVR) based on Sparrow search optimization algorithm (SSA), and obtains the fitting curve through iterative optimization of the hyperparameters. The results show that the prediction accuracy of the least squares support vector regression machine (SSA_LSSVR) based on the sparrow search optimization algorithm is higher, and it can predict the development trend of global power production more accurately.*

Keywords: *Global electricity generation, Sparrow search optimization algorithm, Least squares support vector machine, Time series prediction*

1. Introduction

As we all know, the social development of a country cannot be far away from electricity, which is the material basis of production and life in modern society, and the international trade with electricity has brought more development opportunities to many countries. With the gradual advancement of industrialization in various countries, human demand for electricity will inevitably increase year by year. Whether the power industry structure and development strategy can be adjusted while developing may indirectly affect the development of a country. In the past, most prediction methods used neural network models, such as the BP neural network (BPNN) [1], Radial basis function neural network (RBFNN) [2], long-term and short-term memory artificial neural network (LSTM), etc.

However, because most of these methods require a large amount of historical data, and the actual data provided is very limited, there are certain limitations [3, 4]. In recent years, scholars in our country and abroad have conducted many studies on the regression and classification of vector machines [5, 6]. The research shows that support vector machines have great advantages in solving practical problems such as high dimensions, small samples, nonlinearity, and local extreme value, so they can replace the artificial neural network algorithm for prediction in certain cases. From what has been discussed above, the least squares support vector regression machine (SSA-LSSVR) optimized by the sparrow algorithm is proposed in this paper to predict the global electricity generation, and finally to predict the future trend with high accuracy.

2. Least squares support vector regression

Support vector machine (SVM) is developed as a kind of an intelligent algorithm based on classification or regression at the end of the last century. Support vector machine (SVM) is often found that sometimes than other widely used machine learning algorithm of neural network (such as above) has a better prediction result and be used. It is mainly by looking for both that can meet the requirements of classification and can guarantee the classification accuracy of the optimal hyperplane. It was initially used to solve the linearly separable problem of dichotomies. Then in the process of gradual development, through the application of kernel function, the linear non-separable problem of low dimensional space is mapped to the linear separable problem of high dimensional space, and the linear non-separable problem is solved. Later, by improving the traditional SVM [7], Suykens et al. developed the least squares support vector machine (LS-SVM) [8]. Compared with the standard SVM, LS-SVM uses the linear least squares method to replace the traditional quadratic programming method to solve the loss function, which greatly simplifies the standard SVM.

ISSN 2616-5775 Vol. 4, Issue 5: 1-7, DOI: 10.25236/AJCIS.2021.040501

It has a better adaptability and inherits good generalization ability, which greatly promotes the application of LS-SVM in image recognition and regression analysis [9-10].LS-SVM is described below.

Consider a given training set$\{x_i, y_i\}, \emptyset \in R^2, i = 1,2, \dots N$, First, the input data $x_i$and the output data y_i, We can build regression models by constructing the following nonlinear mapping functions:

$$y = w^t \phi(x) + b \tag{1}$$

Where, w is the weight vector, b is the bias term. Of course, as one of SVM, which must minimize the cost function C containing the penalty regression error:

$$minC(w, e) = \frac{1}{2} w^t w + \frac{1}{2} \gamma \sum_{i=1}^{N} e_i^2 \tag{2}$$

And it's constrained by the following equations

$$y = w^t \phi(x_i) + b + e_i, i = 1,2,3, \dots, N \tag{3}$$

The first part of this cost function is weight attenuation, which is used to gauge weight size and penalize large weights. Because of this regularization, the weights converge to similar values. Large weights can worsen the generalization ability of LS-SVM because they lead to excessive variance. The second part (2) is the regression error of all the training data. Parameter C must be given by the user through optimization in advance. Compared with the first part, this part gives the relative weight. The third part (3) provides the limitation by giving the definition of regression error. To solve this optimization problem, the Lagrange function is constructed as:

$$L(w, b, e, a) = \frac{1}{2} \parallel w \parallel^2 + \gamma \sum_{i=1}^{N} e_i^2 - \sum_{i=1}^{N} \alpha_i \{ w^t \phi(x_i) + b + e_i - y_i \} \tag{4}$$

Where α_i are Lagrange multipliers.

The above solution of (4) can be obtained by partially differentiating with respect to b, w, e_i, α_i:

$$\begin{cases} \begin{cases} \frac{\partial L}{\partial w} = 0 \to w = \sum_{i=1}^{N} \alpha_i \phi(x_i) \\ \frac{\partial L}{\partial b} = 0 \to \sum_{i=1}^{N} \alpha_i = 0 \end{cases} \\ \begin{cases} \frac{\partial L}{\partial e_i} = 0 \to \alpha_i = \gamma e_i, i = 1,2 \dots, N \\ \frac{\partial L}{\partial \alpha_i} = 0 \to w^t \phi(x_i) + b + e_i - y_i = \gamma e_i, i = 1,2, \dots, N \end{cases} \end{cases} \tag{5}$$

And then:

$$w = \sum_{i=1}^{N} \alpha_i \phi(x_i) = \sum_{i=1}^{N} \gamma e_i \phi(x_i) \tag{6}$$

Where a positive definite kernel is used as follows:

$$K(x_i, x_j) = \phi(x_i)^t \phi(x_j) \tag{7}$$

The most important result is that the weights (w) be written as linear combinations of the Lagrange multipliers with the corresponding data training (x_i), Putting the result of (6) into (1), The following formula is obtained:

$$y = \sum_{i=1}^{N} \alpha_i \phi(x_i)^t \phi(x) + b \tag{8}$$

For which $y_i$can be evaluated it is:

$$y_i = \sum_{i=1}^{N} \alpha_i \phi(x_i)^t \phi(x_j) + b \tag{9}$$

The vector follows from solving a set of linear equations:

$$A \begin{bmatrix} \alpha \\ b \end{bmatrix} = \begin{bmatrix} y \\ 0 \end{bmatrix} \tag{10}$$

Where A is a square matrix given by:

$$A = \begin{bmatrix} K + \frac{1}{\gamma} & 1_N \\ I_N^T & 0 \end{bmatrix} \tag{11}$$

Where K denotes the kernel matrix with ijth element in (7) and I denotes the identity matrix $N \times N$, $1_N = [1\ 1\ 1 \dots 1]^T$T Hence, the solution is given by:

ISSN 2616-5775 Vol. 4, Issue 5: 1-7, DOI: 10.25236/AJCIS.2021.040501

$$\begin{bmatrix} \alpha \\ b \end{bmatrix} = A^{-1} \begin{bmatrix} y \\ 0 \end{bmatrix} \quad (12)$$

From (11) to (12), it can be seen that generally, all Lagrange multipliers (support vectors) are non-zero, which means that all training objects contribute to the solution. Compared to standard SVM, LS-SVR solutions are usually not sparse. However, sparse solutions can easily be achieved by pruning or reducing techniques. Depending on the number of training data sets, either a direct solver or an iterative solver such as the conjugate gradient method (for large data sets) can be used, in both cases using numerically reliable methods. In applications involving nonlinear regression, it is enough to change the inner product of $\phi(x_i) \cdot \phi(x_j)$ (9) by a kernel function and the ijth element of the matrix K is equal to (7). If this kernel function satisfies Mercer's case, the kernel implicitly determines both a nonlinear mapping,$x \rightarrow \phi(x)$and the corresponding inner product $\phi(x_i)^t \cdot \phi(x_j)$.And then following nonlinear regression function can be obtained:

$$y = \sum_{i=1}^{N} \alpha_i K(x_i, x) + b \quad (13)$$

For a pointx_i j to be evaluated it is:

$$y_j = \sum_{i=1}^{N} \alpha_i K(x_i, x_j) + b \quad (14)$$

Any function $K(x_i, x_j)$ satisfying Mercer's condition can be used as a kernel function. Typical kernel functions include linear function, polynomial function, radial basis function (RBF), sigmoid function, and so on. Among them, the radial basis function requires fewer parameters and has excellent performance, so RBF is an effective choice of the kernel function, so we will take(15)RBF as the kernel function:

$$K(x, x_i) = exp\left(-\frac{\|x-x_i\|^2}{2\sigma^2}\right) \quad (15)$$

Therefore, compared with the standard SVR, the LS-SVM regression model only requires the user to appropriately select the two main hyperparameters C and σ^2 in advance. The selection of hyperparameters plays an important role in the performance of LS-SVM. Among the alternatives, it is best to use cross-validation. Based on this idea, some traditional methods are used to obtain the optimal hyperparameters of the regression model. For example, genetic algorithm, simulated annealing algorithm, and particle swarm optimization algorithm, and other evolutionary methods, are one of the most widely used methods, but this paper uses a relatively new Sparrow Search optimization algorithm.

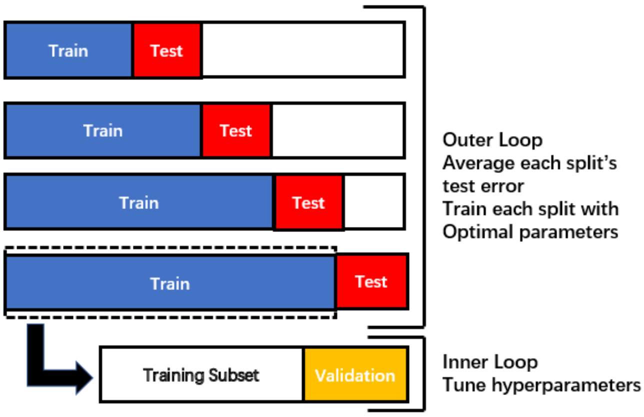

Figure 1: Nested Cross-Validation

3. Sparrow search optimization algorithm

As we all know, sparrows are omnivorous migratory birds widely distributed in the world. For a long time, their wisdom is reflected in their coexistence with human beings. By observing the habits of sparrows, we can first divide them into two classes: producers and scroungers.

The scroungers get their food from producers while producers actively search for food sources. Long observations of sparrows show that they have a flexible strategy for acquiring food, shifting between the roles of producer and scrounger.

To simplify the analysis of this process, we will idealize it and follow these rules:

(1) The producer usually has a high level of energy reserves and provides a place or direction for all scroungers. It is responsible for identifying areas where rich food sources can be found. The level of energy reserve depends on the assessment of the individual fitness value.

(2) Each sparrow could become a producer if it sought a better source of food, but the proportion of producers and scroungers did not change across the population.

(3) As soon as the sparrows detect a predator, they begin to call as an alarm signal. When the alarm value is greater than the safety threshold, the producer needs to direct all scroungers to the safety zone.

(4) When the sparrows at the edge of the group realized the danger, they quickly moved to the safe area to get a better position, while the sparrows in the middle of the group moved randomly to get close to other sparrows.

(5) Sparrows with higher energy as producers. Some hungry scroungers are more likely to fly elsewhere for food in order to get more energy.

(6) The scroungers go in search of food after the producer who can provide it best. At the same time, some scroungers may constantly monitor producers and compete for food to increase their own rate of predation.

In the modeling process of sparrow search optimization algorithm, we tend to virtual the process of the sparrow to find food:

$$X = \begin{bmatrix} x_{1,1} & x_{1,2} & x_{1,3} & x_{1,4} & \cdots & \cdots & x_{1,d} \\ x_{2,1} & x_{2,2} & x_{2,3} & x_{2,4} & \cdots & \cdots & x_{2,d} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{n,1} & x_{n,2} & x_{n,3} & x_{n,4} & \cdots & \cdots & x_{n,d} \end{bmatrix} \tag{16}$$

Where n the number of sparrows, d shows the dimension of the variables to be optimized. the fitness value of all sparrows can be expressed by the following vector:

$$Fx = \begin{bmatrix} f([x_{1,1} & x_{1,2} & \cdots & \cdots & x_{1,d}]) \\ f([x_{2,1} & x_{2,2} & \cdots & \cdots & x_{2,d}]) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ f([x_{n,1} & x_{n,2} & \cdots & \cdots & x_{n,d}]) \end{bmatrix} \tag{17}$$

Where n represents the number of sparrows, and the value of each row in Fx represents the fitness value of the individual. In SSA, producers with higher fitness values were given priority in obtaining food during the search process. In addition, because producers are responsible for finding food and directing the movement of the entire population. As a result, producers can find food in a much wider range than scroungers.

According to the above rules, in each iteration.

The location of the producer is updated as follows:

$$X_{i,j}^{t+1} = \begin{cases} X_{i,j}^{t} \cdot \exp\left(\frac{-i}{\alpha \cdot iter_{max}}\right) & if\ R_2 < ST \\ X_{i,j}^{t} + Q \cdot L & if\ R_2 \geq ST \end{cases} \tag{18}$$

Where t represents the number of current iterations, $j = 1,2,3, \dots, d$. $X_{i,j}^{t}$ represents the value of the jth dimension of the ith sparrow at iterations t. $iter_{max}$ is a constant with the largest number of iterations. $\alpha \in (0,1]$ is a random number. R_2 ($R_2 \in [0,1]$) and ST ($ST \in [0.5,1.0]$) represent the alarm value and the safety threshold, respectively. Q is a random number that obeys normal distribution.L is a matrix of $1 \times d$, which each element inside is 1. By comparing R_2 with ST, we can judge whether it is disturbed by predators.

The position of scroungers has been updated below:

$$X_{i,j}^{t+1} = \begin{cases} Q \cdot \exp\left(\frac{X_{worst}^{t} - X_{i,j}^{t}}{i^2}\right) if\ i > \frac{n}{2} \\ X_p^{t+1} + \left|X_{i,j}^{t} - X_p^{t+1}\right| \cdot A^{+} \cdot L \quad if\ i \leq \frac{n}{2} \end{cases} \tag{19}$$

X_p is the best position occupied by the producer. X_{worst} indicates the globally worst location.A represents a matrix of $1 \times d$ for which each element inside is randomly assigned 1or-1, and $A^{+} = A^{T}(AA^{T})^{-1}$。 When $i > \frac{n}{2}$, it suggests that the ith scrounger with the worse fitness value is most likely to

ISSN 2616-5775 Vol. 4, Issue 5: 1-7, DOI: 10.25236/AJCIS.2021.040501

be starving. In the simulation, we assumed that these sparrows were 10 to 20 percent of the population aware of the danger. We make the initial position of these sparrows in the population random.

Based on the above rules, we can build the following mathematical model:

$$X_{i,j}^{t+1} = \begin{cases} X_{i,j}^{t} + K \cdot \left(\frac{\left| X_{i,j}^{t} - X_{worst}^{t} \right|}{(f_i - f_w) + \varepsilon} \right) & if\ i > \frac{n}{2} \\ X_{best}^{t} + \beta \cdot \left| X_{i,j}^{t} - X_{best}^{t} \right| & if\ f_i > f_g \end{cases} \quad (20)$$

Where$X_{best}$is the current global best position. As the step size control parameter,β is a normally distributed random number with the mean value of 0 and a variance of 1.K $\in$[-1,1] is a random number, which is the direction of the sparrow's movement and the step size control coefficient. Here f_i is the fitness value of the sparrow at present. f_g and f_w are now the global best and worst fitness values, respectively.ε is the smallest constant to avoid zero division error. When$f_i > f_g$, indicates that sparrows are located at the edge of the entire population. X_{best} means the central position of the whole group, the safest position. Based on the above algorithms, we can optimize the two main hyperparameters γ and σ^2 of the LS-SVM regression model by using the Sparrow optimization algorithm. In solving the hyperparameter selection, each sparrow represents a potential solution consisting of the vector $d = (\gamma, \sigma^2)$. Hyperparametric optimality is measured by defining fitness functions related to the optimization problem under consideration. During the training and testing of LS-SVM, the goal is to improve the generalization performance of the regression model, that is, to minimize the error between the true value and the predicted value of the test sample. Therefore, the fitness function can be defined as:

$$Fitness = \frac{1}{n}\sum_{i=1}^{n}\sqrt{\frac{1}{m}\sum_{j=1}^{m}\left(f\left(x_{ij}\right) - y_{ij}\right)^2} \quad (21)$$

The modeling programming environment used in this paper is MATLAB R2018a, and the data is from BP-STATs-Review-2021-ALL-Data.At the same time, we took the data of Electricity Generation from the above literature to carry out prediction study, and the following are the specific implementation process:

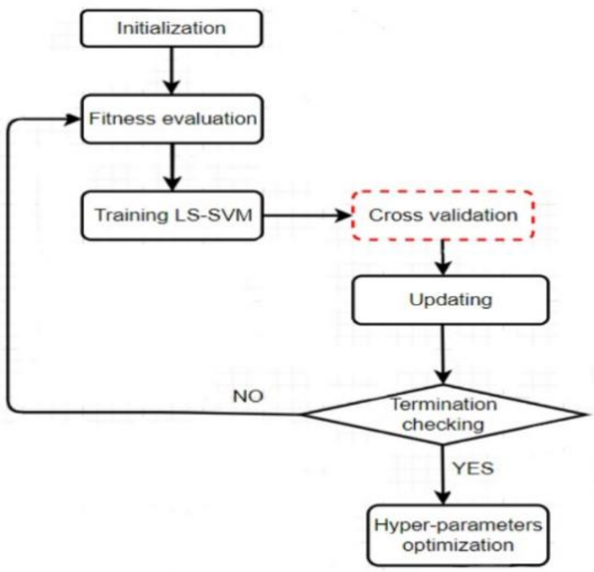

Figure 2: Flowchart of SSA for hyper-parameter optimization

Then, the results are shown as follows:

ISSN 2616-5775 Vol. 4, Issue 5: 1-7, DOI: 10.25236/AJCIS.2021.040501

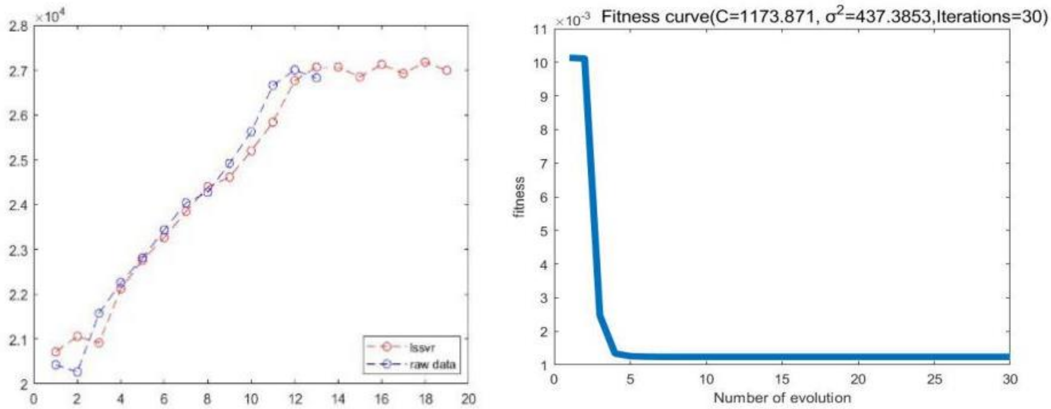

Figure 3: The prediction results based on the optimization algorithm

After that, we conducted correlation analysis on the above regression curves:

We select $x_1, \dots, x_n$ as training data, $f(x_1), \dots, f(x_n)$ as the predictive value of LS-SVM, $y_1, \dots, y_n$ as the raw data. Finally, we evaluated this model by MAPE and R^2:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{f(x_1)-y_i}{y_i}\right| \times 100\% \tag{22}$$

$$R^2 = \frac{(n\sum_{i=1}^{n} f(x_i)y_i - \sum_{i=1}^{n} f(x_i)\sum_{i=1}^{n} y_i)^2}{(n\sum_{i=1}^{n} f(x_i)^2 - (\sum_{i=1}^{n} f(x_i))^2)(n\sum_{i=1}^{n} y_i^2 - (\sum_{i=1}^{n} y_i)^2)} \tag{23}$$

	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026
True value	*20422*	*20265*	*20571*	*22257*	*22806*	*23435*	*24032*	*24271*	*24915*	*25624*	*26659*	*27001*	*26823*						
Fitting value	*20712*	*21059*	*20914*	*22122*	*22756*	*23264*	*23843*	*24390*	*24609*	*25196*	*25837*	*26762*	*27065*						
Predicted value														*27065*	*26847*	*27122*	*26924*	*27175*	*26994*

Figure 4: The specific value of regression curve

R^2	$Adjusted - R^2$	PCCs	MAPE
0.970096	0.967377	0.984934	0.014609511603398

Figure 5: Fitting curve correlation analysis parameters

4. Conclusion

In this paper, a new prediction model, the least squares support vector regression model (LS-SVM) based on sparrow optimization algorithm (SSA), is studied to predict global power production by optimizing the hyperparameters.

Compared with the standard SVM, LS-SVM simplifies the standard SVM to a large extent and replaces the traditional quadratic programming method to solve the loss function by applying the linear least squares method.

The super parameter optimization based on the SSA algorithm has an excellent fitting effect without considering other objective and small probability events. The results show that the annual global power output will fluctuate and rise, reaching a recent peak in 2025.

References

[1] H.S. Hippert, C.E. Pedreira, R.C. Souza, Neural networks for short-term load forecasting: A review

ISSN 2616-5775 Vol. 4, Issue 5: 1-7, DOI: 10.25236/AJCIS.2021.040501

and evaluation, IEEE Trans. Power Syst. 16 (1) (2001) 44–55.
[2] F.J. Chang, Y.C. Chen, Estuary water-stage forecasting by using radial basis function neural network, J. Hydrol. 270 (1/2) (2003) 158–166.
[3] M.H. Wang, C.P. Hung, Novel grey model for the prediction of trend of dissolved gases in oil-filled power apparatus, Electr. Power Syst. Res. 67 (1) (2003) 53–58.
[4] S.W. Fei, Y. Sun, Forecasting dissolved gases content in power transformer oil based on support vector machine with genetic algorithm, Electr. Power Syst. Res. 78 (3) (2008) 507–514.
[5] F.E.H. Tay, L.J. Cao, Application of support vector machines in financial time series forecasting, Omega-Int. J. Manage. Sci. 29 (4) (2001) 309–317.
[6] Q. Wu, The forecasting model based on wavelet v-support vector machine, Expert Syst. Appl. 36 (4) (2009) 7604–7610.
[7] V.N. Vapnik, the Nature of Statistical Learning Theory, Springer-Verlag, New York, 1998.
[8] J.A.K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, J. Vandewalle, Least Squares Support Vector Machines, World Scientific, Singapore, 2002.
[9] T. Van Gestel, J.A.K. Suykens, D.E. Baestaens, Financial time series prediction using least squares support vector machines within the evidence framework, IEEE Trans. Neural Netw. 12 (4) (2001) 809–821.
[10] Z. Yang, X.S. Gu, X.Y. Liang, L.C. Ling, Genetic algorithm-least squares support vector regression based predicting and optimizing model on carbon fiber composite integrated conductivity, Mater. Des. 31 (3) (2010) 1042–1049.