

Research on Short-Term Load Forecasting Based on Improved Support Vector Regression

Baoyi Wang^a, Tianyang Han^b, Shaomin Zhang^c

School of Control and Computer Engineering, North China Electric Power University, Baoding 071003, Hebei Province, China

^aemail:wangbaoyiqj@126.com, ^bemail:407927192@qq.com, ^cemail:zhangshaomin@126.com

Keywords: artificial bee colony algorithm; support vector machine; micro grid; load forecasting

Abstract. Firstly, according to the disadvantages of the artificial bee colony algorithm which is easy to fall into the local optimum and the convergence rate is slow, this paper introduce the current optimal food source and inertia weight function to improve the food source updating method. Then according to the parameter optimization of support vector regression, this paper transform it into a combinatorial optimization problem, and use the improved artificial bee colony algorithm to solve the optimization problem, and then establish the prediction model of artificial bee colony algorithm to optimize the SVR. Taking the short-term load forecasting data of micro-grid as an example, the prediction results of the model are compared and analyzed. The results show that the model has the best forecasting effect and the shortest running time, and has better learning ability and generalization ability.

Introduction

Support Vector Machine (SVM), based on statistical learning theory, was first developed by Vapnik in 1995 [1-2]. When applied on the small sample set of nonlinear and high-dimension pattern recognition problems, it showed many special advantages. For example, it can avoid the curse of dimensionality and is easy to compute [2]. Till now, SVM has become a hot topic of intensive study due to its successful application in classification and regression tasks.

Karaboga [6] firstly proposed the ABC algorithm in optimizing numerical functions, and after that, several studies proved that ABC shows better performance than GA and PSO algorithm [7,8]. In different optimization problems and application platforms, many studies suggested adopting ABC methodology in both standard and improved forms [9,10].

This paper aimed at the shortage of artificial bee colony algorithm, improved its own search mechanism, increase the current optimal food source and the inertia weight function, optimize the food source update method. Then use the improved artificial bee colony algorithm to optimize the parameters of support vector regression finally this model is applied to the micro grid short-term load forecasting and to compare the results of that.

Support Vector Regression Machine Theory

Based on the structure risk minimization principle, SVM method seeks to minimize an upper bound of generalization error instead of the empirical error as in other neural networks. Additionally, the SVM models generate the regression function by applying a set of high-dimensional linear functions. The SVM regression function is formulated as follows:

$$f(x) = w \cdot \Phi(x) + b \quad (1)$$

where $\Phi(x)$ is called the feature, which is nonlinear and mapped from the input space R^n . $f(x)$ is the target output value we want to estimate. The coefficients w and b estimated by minimizing:

$$R(f) = \frac{1}{2} \|w\|^2 + C \frac{1}{l} \sum_{i=1}^l L_{\varepsilon}(y_i, f(x_i)) \quad (2)$$

where:

$$L_{\varepsilon}(y_i, f(x_i)) = \begin{cases} |y_i - f(x_i)| - \varepsilon, & |y_i - f(x_i)| \geq \varepsilon, \\ 0, & \text{others} \end{cases} \quad (3)$$

Eq.(2) is the risk function consisting of the empirical error and a regularization term which is derived from the SRM principle. The term $\frac{1}{l} \sum_{i=1}^l L_{\varepsilon}(y_i, f(x_i))$ in Eq.(2) is the empirical error (risk), which is measured by the ε -insensitive loss function (ε -insensitive tube) given by Eq.(3); in the meanwhile, the term $\frac{1}{2} \|w\|^2$ is the regularization term. The constant $C > 0$ is taken as the regularized constant that determines the trade-off between the empirical error (risk) and the regularization term. Increasing the value of C will add importance to the empirical risk in the risk function. ε is called the tube size of the loss function and it is equivalent to the accuracy approximation placed on the training data points. Both C and ε are user-prescribed parameters.

Then, the slack variables ξ_i and ξ_i^* , which represent the distance from the actual values to the corresponding boundary values of ε -insensitive tube are introduced. With these slack variables, Eq.(3) can be transformed to the following constrained formation:

Minimize:

$$\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) \quad s. t. \begin{cases} y_i - w \cdot \Phi(x_i) - b \leq \varepsilon + \xi_i \\ w \cdot \Phi(x_i) + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \quad (i = 1, 2, \dots, l) \end{cases} \quad (4)$$

Finally, by introducing the Lagrangian multipliers and maximizing the dual function of Eq.(4), Eq.(4) can be changed to the following form:

$$\max -\frac{1}{2} \sum_{i,j=1}^l (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) K(x_i, x_j) - \sum_{i=1}^l \alpha_i (\varepsilon - y_i) - \sum_{i=1}^l \alpha_i^* (\varepsilon + y_i) \quad s. t. \begin{cases} \sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0 \\ 0 \leq \alpha_i, \alpha_i^* \leq C \quad (i = 1, 2, \dots, l) \end{cases} \quad (5)$$

In Eq.(5), α_i and α_i^* are called Lagrangian multipliers and they satisfy the equalities $\alpha_i * \alpha_i^* = 0$, the general form of the regression estimation function can be written as:

$$f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) K(x_i, x_j) + b \quad (6)$$

In this equation, $K(x_i, x_j)$ is called the kernel function. Kernel function is a symmetric function $K(x_i, x_j) = \Phi(x_i) \Phi(x_j)$ satisfying Mercer's conditions, when the given problem is a nonlinear problem in the primal space, we may map the sample points with mapping into a high-dimensional feature space where linear problem can be performed. Linear, Polynomial, Radial Basis Function (RBF) and sigmoid are four main kernel functions in use. In most of the real applications, the SVR employs RBF kernel function to estimate the nonlinear behavior of the given data set because RBF kernels tend to give good performance under general smoothness assumptions.

The Basic Theory Of Artificial Swarm Algorithm.

The basic ABC algorithm can be divided into three phases—employed bee phase, onlooker bee phase, and scout bee phase. At first, the food sources need to be selected. A set of possible solution to a problem can be taken as food sources. So, if we have a D -dimensional problem (where $i=1, 2, \dots, D$), then the decision variables (x_i) act as the members of a single food sources. The values of these decision variables are randomly generated within the variable bounds if they have any. The employed bee tests the fitness values of predefined number of food sources and records the information about them. The quality of each source is the fitness value of the objective function considered to minimize or maximize according to service criteria of the reservoir. Employed bees select a food source and randomly choose a candidate solution within the source to update it, providing new values by using Eq.(7):

$$v_{ij} = x_{ij} + r_{ij}(x_{ij} - x_{kj}) \quad (7)$$

Here, x_{ij} represents the current candidate solution of a source, and x_{kj} represents other randomly chosen solution but must be taken from a different neighbor source. r_{ij} is taken as a random number between $[-1, 1]$. In case of any violation of variable limits (after using Eq.(7)), the updated food sources are needed to crop to its allowable ranges. The qualities (fitness values) of the new food

sources are checked again with the previously stored optimum solution, and the decision for the survival of a particular source solely depends on the fitness of the solutions.

The onlookers begin to work with the information provided by the employed bees about the food sources. Onlookers select the sources according to their probability of being selected. The probability of each source depends on their fitness value, and for total N number of sources, it can be expressed as Eq.(8):

$$p_i = \frac{fit_i}{\sum_{n=1}^{SN} fit_n} \quad (8)$$

In Eq.(8), the probability (p_i) of any i th source is the ratio of its individual fitness to the sum of the fitness of all sources. After selection, onlookers use Eq.(7) to maintain the diversion of the sources. The scout bees create new sources randomly to replace the abandoned food sources (the source posing weak fitness value).

In the bee colony algorithm, the parameter *limit* is used to save the update times of the solution. If after *limit* cycles, a solution does not change, then it has been trapped in a local optimal solution, the solution will be removed, corresponding to the solution of the lead bee has become a scout bee. The elimination of the solution is x_i the reconnaissance bee in accordance with the Eq.(9) randomly generated a new solution to replace x_i .

$$x_{ij} = \min_j x_{ij} + rand(0,1) \left(\max_j x_{ij} - \min_j x_{ij} \right) \quad (9)$$

In the formula, $j \in \{1, 2, \dots, D\}$, x_{ij} is the first j element of vector x_i .

The Improvement Of Artificial Bee Colony Algorithm.

In basic artificial bee colony algorithm, employed bee and onlooker bee are in accordance with the Eq.(7) search nearby the food source, this update method is only to find a new solution near the optimal solution., although the search method has strong local search ability, but the lack of global search ability, the individual's search space is relatively small, and other individuals do not exchange information, reducing the probability of success of the updated solution. In order to overcome these shortcomings, in the original basis to add the current global optimal solution, the introduction of two kinds of weight function, employed bee and onlooker bee's new food sources were updated according to the Eq.(10,11).

$$v_{ij} = x_{ij} + r_{ij} w_i^1 (x_{ij} - x_{kj}) \quad (10)$$

$$v_{ij} = x_{ij} + r_{ij} w_i^1 (x_{ij} - x_{kj}) + w_i^2 (x_{bj} - x_{ij}) \quad (11)$$

Among,

$$w_i^1 = 1 - \left(w_{max} - (w_{max} - w_{min}) \left(\frac{k}{T_{max}} \right) \right), \quad w_i^2 = w_{max} (w_{max} - w_{min}) \left(\frac{T_{max} - k}{T_{max}} \right) \quad (12)$$

In this paper, we use different methods to update employed bee and onlooker bee. Employed bees are associated with a specific source of food, the food source is they are currently collecting nectar, to share with other bees. In order to share information better, increase an increasing weight function w_i^1 . With the increase of the number of iterations the function can be adaptive to the control of the neighborhood range, which is conducive to improving the accuracy and convergence of the swarm optimization algorithm. A new method for the production of employed bee was obtained Eq.(10). Onlooker bee according to Eq.(11) to update. The update method is to add the current optimal solution x_{bj} and decreasing weight function w_i^2 , the current optimal solution x_{bj} can expand the scope of the neighborhood search and enhance the exchange of information, which is beneficial to improve the search efficiency. And w_i^2 is to control the food source inherit previous information, when the inertia weight is small for the local search, the inertia weight is larger for global search. With the iteration of the algorithm, the initial larger weight can keep the algorithm to maintain a strong global search performance, and to the later iteration, the smaller weight can make the algorithm local search more accurate.

Experiment And Analysis

This paper selects the micro grid demonstration project in July 2011, a week before 168h load data as the training set, July 8th 24h load data for test set. The collection includes daily load power, real-time temperature, weather and day type. Load power, real-time temperature recorded once every 15 minutes, the weather and day type record once a day, from zero to twenty-four every 15 minutes a point, a total of 96 values.

Model Parameter Set. In order to better compare the performance of the algorithm, the common parameters of ACO-SVR, PSO-SVR, ABC-SVR and the improved ABC-SVR are all set: the evolution algebra is 100, the population size is 20, the loss function ε value is in the range of (0,1], the kernel function related parameter σ value is in the range of (0,1], the scope of penalty parameter C value is in the range of (0,10). The other algorithm parameters are set according to their optimal performance.

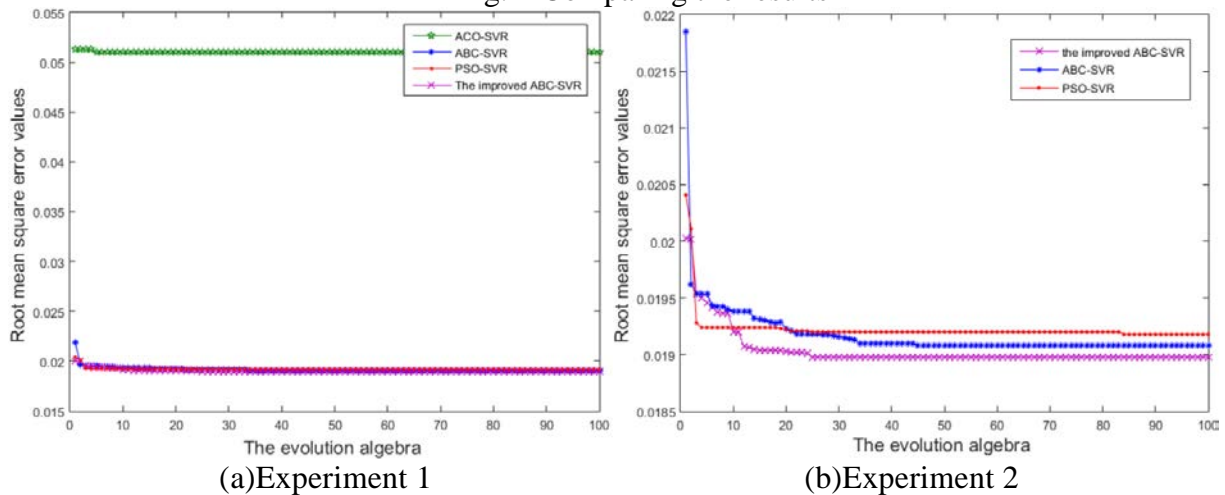
Instance Analysis Results. Respectively using ACO-SVR, PSO-SVR, ABC-SVR and the improved ABC-SVR model to forecast the short-term load of micro grid, get the loss function ε , the parameters of the kernel function σ and penalty factor C value.

Table. 1 The parameter optimization results

Model	ε	σ	C
ACO-SVR	0.5783	0.8343	9.5486
PSO-SVR	0.2263	0.7670	3.9751
ABC-SVR	0.2260	0.9011	2.8473
The improved ABC-SVR	0.2891	0.8780	1.6287

As can be seen from the above table, the parameters of the support vector regression machine are optimized by using different intelligent algorithms, and the final parameters are different. Each intelligent algorithm in the parameter optimization process is shown in Fig. 1.

Fig. 1 Comparing the results



From Fig. 1(a) we can see that the error value of the ACO-SVR is the highest, that is, the prediction effect is the worst, and the other models were not so far different, in order to make clear that which is the optimal prediction model, from Fig. 1(b) we can clearly understand that the root mean square error minimum namely the highest prediction accuracy was the improved ABC-SVR model, followed by ABC-SVR and PSO-SVR. Look from the convergence rate, PSO-SVR, the fastest convergence, slow convergence to the late; The improved ABC-SVR and ABC-SVR in the starting process is similar, but in the later stage, the improved ABC-SVR was significantly more stable than ABC-SVR, and the effect was better than the ABC-SVR model, which shows that the improved model can improve the convergence speed and prediction accuracy.

In order to compare the performance of each forecast model more clearly, using mean absolute percentage error (MAPE) and root mean square error (RMSE) two indicators to measure the pros

and cons of the prediction model, the smaller the MAPE and the RMSE, the more accurate the prediction results are. The specific values are shown in Table. 2.

Table. 2The model prediction results contrast

Model	MAPE	RMSE	TIME
ACO-SVR	1.0703	37.2346	71.1547
PSO-SVR	0.5411	22.2959	13.7198
ABC-SVR	0.5424	22.2387	15.1274
The improved ABC-SVR	0.5268	22.1782	8.8098

From the table, we can see the improved ABC-SVR forecasting model is the most optimal prediction effect and shortest in current models, which shows that the model can improve the running efficiency and has good performance.

Summary

(1)The optimization of parameters of SVR determines its performance, which affects its generalization ability and learning ability. Artificial bee colony algorithm is a new swarm intelligence evolutionary algorithm, which has strong global search and local search performance, however, it still has the disadvantage of slow convergence speed, so, this paper improves the updating method based on the original algorithm. The convergence rate and prediction accuracy of the algorithm are improved to a certain extent. Then the improved algorithm is applied to optimize the parameters of support vector machine, and an improved ABC-SVR prediction model is proposed. The results show that the proposed algorithm can solve the problem of parameter optimization of support vector regression machine well.

(2)The improved ABC-SVR model is used to predict the short-term load of micro-grid, and comparative analysis with PSO-SVR, ACO-SVR and ABC-SVR model. The improved ABC-SVR improves the prediction accuracy and running efficiency. It shows that the model has good generalization ability and generalization ability. From the prediction results, this model is suitable for short-term load forecasting of microgrid. The parameter optimization of support vector regression machine is a far-reaching and influential problem, so the next will focus on its nature for more in-depth research and analysis.

Acknowledgement

In this paper, the research was sponsored by the National Natural Science Foundation of China (Project No.) and Scientific Research Project of Hebei Province (Project No.).

References

- [1] Vapnik V.The nature of statistical learning theory[M]. New York:wiley,1998.
- [2] Olivier Chapelle, Vladimir Vapnik, Olivier Bousquet, Sayan Mukherjee. Choosing Multiple Parameters for Support Vector Machines[J]. Machine Learning,2002,46:131-159.
- [3] Mingyuan Zhao, Yong Tang, Chong Fu, Mingtian Zhou. Based on the feature of chromosomes genetic algorithm of support vector machines feature selection and parameters optimization [J]. Control and Decision, 2010,08:1133-1138.
- [4]Ning Lu, Benling Wu, Ying Liu. The SVM model based on adaptive particle swarm optimization in the application of the load forecasting [J]. Power System Protection and Control,2011,15:43-46.
- [5]Wen Long, Xi-Ming Liang, Zuqiang Long, Zhaohui Li. Based on improved ant colony algorithm to optimize parameters of LSSVM short-term load forecasting [J]. Journal of Central South University(Science and Technology),2011,11:3408-3414.

- [6]Basturk B,Karaboga D.An artificial bee colony (ABC) algorithm for numeric function optimization. Proceedings of IEEE Swarm Intelligence Symposium Indianapolis. Indianapolis, USA: 2006: 651-656.
- [7]Karaboga D, Basturk B.A powerful and efficient algorithm for numerical function optimization: artificial bee colony(ABC) algorithm.Journal of Global Optimization, 2007,39(3):459-471.
- [8]Karaboga D, Basturk B. On the performance of artificial bee colony(ABC) algorithm. Applied Soft Computing,2008,8(1): 687-697.
- [9] Xiaojun Bi, Yanjiao Wang. Niche artificial bee colony algorithm for multimodal function optimization [J]. Systems Engineering and Electronics,2011,11: 2564-2568.
- [10]Li Fu, Jun Luo. Artificial bee colony algorithm with tracking search and immune selection [J]. Pattern Recognition and Artificial Intelligence,2013,07:688- 694.
- [11]Jianxia Yin, Hongyun Meng. Artificial colony algorithm with chaos differential evolution search [J]. Computer Engineering and Applications,2011,29:27-30.
- [12]Shujuan Yu, Jun Zhang, Lei Yang. Blind detection algorithm based on improved ant colony optimization [J]. Computer Technology and Development,2013,11:74-76+81.
- [13]Ming Chen, Yanmin Liu. Based on improved particle swarm algorithm of adaptive rejection factor [J]. Journal of Computer Applications,2013,08:2269-2272.