SVM-based Implicit Stochastic Scheduling Mode for Cascade Hydropower Stations

Wang Qinhui¹, Li Jidong^{1,a}, Chen Shijun^{2,3}, Wang xiaoyi¹

¹ College of Water Conservancy and Hydropower Engineering, Sichuan Agricultural University, Ya'an 625014, China;

² College of Water Resources and Hydropower, Sichuan University, Chengdu 610065, China;

³ Business School, Sichuan University, Chengdu 610065, China

Abstract. In view of the uncertainty of inflow of cascade reservoirs, an implicit stochastic joint scheduling function model for cascade hydropower stations based on support vector machine (SVM) was established. Taking the cascade hydropower station on the lower reaches of the Yalong River as an example, according to the long series of optimized scheduling simulation operation data, the Gauss radial basis (BRF) kernel function is utilized by LIBSVM for the scheduling function fitting of the cascade reservoirs in the lower reaches of the Yalong River. Besides, combined with particle swarm optimization algorithm, the support vector machine (SVM) model parameters c (penalty coefficient) and g (relaxation coefficient) were optimized. Eventually, the optimized scheduling function model was used for the cascade hydropower station simulation operation. The results show that compared with the existing scheduling technology, the nonlinear SVM scheduling function is better than the linear regression model, and the effect of the nonlinear SVM scheduling function is equivalent to that of the threshold regression model. Therefore, the SVM-based Implicit Stochastic Scheduling method can provide references for the actual operation of the cascade hydropower station.

1 Introduction

With the slowdown of hydropower development, the optimal operation of cascade reservoirs has attracted more and more attention. At present, there are two modes of optimal scheduling methods for cascade hydropower stations: explicit stochastic optimization scheduling and implicit stochastic optimization scheduling. The idea of implicit stochastic optimization of hydropower station was proposed by American scholars in 1967^[1]. The purpose is to find out the scheduling rules from the optimal scheduling process, and subsequently utilize the optimal scheduling theory for the tools that guide the actual operations, such as scheduling function, optimal operation chart, etc.^[2].

Major methods for establishing the cascade reservoir function model include multiple linear regression model, threshold regression model and artificial neural network model^[3]. Research results show that the above-mentioned methods have certain limitations. For example, multiple linear regression models cannot reflect the multidimensional nonlinear relationship among influence factors^[4], and the threshold regression method is limited when the amount of data is small^[5], while the artificial neural network model network structure is difficult to determine^[4]. Support Vector Machine (SVM) is a machine learning classifier based on statistical learning theory. It can be used to solve classification or regression problems^[6]. It avoids the explicit expression of nonlinear mapping because of the

introduction of kernel function. Although its calculation amount is equivalent to the linear regression method, it can reflect the multidimensional nonlinear relationship. Compared with other nonlinear regression methods, the SVM regression algorithm is simple, easy to implement with less complex computation, which enable this algorithm to excel in solving large-scale problems^[4]. Therefore, this investigation intends to solve the problem of implicit stochastic scheduling of cascade reservoirs based on support vector machine technology, so as to provide reference for the actual operation of cascade hydropower stations.

2 Basic theory

2.1 Support Vector Machine

At present, support vector machine classification technology has been widely applied to machine learning, pattern recognition, pattern classification, computer vision, industrial engineering and aerospace^[7]. The basic idea of this technology is to define the optimal linear hyperplane, and summarize the algorithm for finding the optimal linear hyperplane as solving an optimal convex quadratic programming problem^[4]. Take the principle of the solution of the nonlinear regression problem based on support vector machine as an example, the steps can be summarized as follows:

^a Corresponding author: Li Jidong (1990-), male, Ph.D., research direction is the optimization of water and electricity management, E-mail: sdj pz@sicau.edu.cn.

(1) For a given sample point that cannot be divided by a hyperplane, the kernel function is used to map the sample points to a high-dimensional space, in other words, the nonlinear problem is converted into a linear problem in a high-dimensional space.

As for the nonlinear support vector machine, the Gauss radial basis kernel function is a common kernel function, and it can be expressed as:

$$K(\chi_i, \chi_j) = \exp(-\gamma \|\chi_i - \chi_j\|^2), \gamma > 0 \quad (1)$$

(2) After the kernel function is converted, the decision function can be expressed as:

$$f(\chi) = w^T \times \varphi(\chi) + b \quad (2)$$

Where w is the weight coefficient, b is the bias phase.

Finding the optimal w and b , and minimizing the confidence interval when the linear decision function remains unchanged, the optimization problem can be expressed as:

Finding the optimal w and b , and minimizing the confidence interval when the linear decision function remains unchanged, the optimization problem can be expressed as:

$$\min_w = \frac{1}{2} w^T w \quad (3)$$

$$b = \frac{1}{N_{NSV}} \left\{ \sum_{0 < a_i < C} \left[y_i - \sum_{x_j \in SV} (a_j - a_j^*) K(x_j, x_i) - \varepsilon \right] + \sum_{0 < a_i^* < C} \left[y_i - \sum_{x_j \in SV} (a_j - a_j^*) K(x_j, x_i) + \varepsilon \right] \right\} \quad (10)$$

Where N_{NSV} is the number of standard support vectors, a_i^* and a_j^* are Lagrangian multipliers, and C is the penalty coefficient, $K(x_j, x_i)$ is a kernel function.

2.2 Particle Swarm Optimization

Particle Swarm Optimization (PSO) is a new type of bionic algorithm proposed by Kennedy and Eberhart in 1995. The algorithm originates from a study on predation behavior of birds. The basic idea of this algorithm is to find the optimal solution based on the information transmission and information sharing between individuals in the group. This algorithm features less adjustment parameters, simple operation, easy implementation and strong versatility [8-10]. At present, there are certain achievements concerning PSO algorithm research and PSO algorithm is widely used for function optimization, constraint optimization, engineering design problems, and power system fields. Therefore, this investigation optimized the main parameters c and g of SVM by using PSO algorithm.

3 Model construction

3.1 Case Analysis

Taking the cascade hydropower station on the lower reaches of the Yalong River as an example, an implicit stochastic optimization scheduling function model was constructed. The Yalong River is the largest tributary of the upper reaches of the Yangtze River and it is rich in hydropower resources. The segment from *Lianghekou* to

$$y_i - w \times \varphi(x_i) - b \leq \varepsilon \quad (4)$$

$$w \times \varphi(x_i) + b - y_i \leq \varepsilon \quad (5)$$

Where ε is the error.

When the constraint cannot be achieved, the slack variables ξ_i and ξ_i^* is introduced and the optimization problem can be expressed as:

$$\min_{w, \xi, \xi^*} = \frac{1}{2} w^T w + C \sum_{i=1}^n (\xi_i + \xi_i^*) \quad (6)$$

$$y_i - w \times \varphi(x_i) - b \leq \varepsilon + \xi_i \quad (7)$$

$$w \times \varphi(x_i) + b - y_i \leq \varepsilon + \xi_i^* \quad (8)$$

Where $\xi_i \geq 0$, $\xi_i^* \geq 0$, $i = 1, 2, \dots, n$.

(3) The convex quadratic programming problem can be solved by using the Lagrange multiplier method.

(4) Through the regression estimation function obtained by learning, the obtained regression estimation function can be used for nonlinear classification or regression prediction. The regression estimation function can be expressed as:

$$f(x) = \sum (a_i - a_i^*) K(x_i, x) + b \quad (9)$$

the middle and lower reaches of Jiangkou is listed as a national hydropower base. According to altitude, from high to low, the five hydropower stations along the lower reaches of Yalong River are Jinping Station I (JY), Jinping Station II (JE), Guandi Station (GD), Ertan Station (ET) and Tongzilin Station (TZL). These hydropower stations are operated jointly and thus form the cascade reservoir. The cascade reservoir is mainly responsible for power generation, with no other comprehensive utilization requirements [3]. Because the cascade reservoir scheduling method is affected by the accuracy of runoff prediction and the uncertainty of incoming water, in order to improve the operation level of cascade reservoirs, an implicit stochastic scheduling function model of PSO-SVM was constructed based on SVM method to guide the operation of cascade reservoirs.

3.2 Creation of Sample Data

Among the cascade hydropower stations on the lower reaches of the Yalong River, JY is the leading station with annual regulating capacity. The ET station is regarded as the fourth level, and it has seasonal adjustment capability. JE station, GD station and TZL station only have daily regulating capacity, whose water levels remain steady in long-term operation. Therefore, their regulation of runoff was ignored in this study. In order to facilitate the comparative analysis, the model input and output selected in this study are consistent with the literature [3]. Besides, the water levels at the end of each period of JY and ET are the decision variables. The JY initial water levels of the period, inflow flow of the

JY-ET interval as well as the ET initial water level of the period are independent variables. These variables were regarded as the scheduling function model input.

In this study, the long-range runoff data from November 1953 to October 2012 is analysed. The calculation period is from November to December of the following year, and the calculation cycle is one month. Simulation calculation was performed by using the progressive optimality algorithm (POA). The results were used as data samples for implicit stochastic optimization calculation. The data from November 1953 to October 2002 in the data samples were used as training samples to determine the model parameters, while the data from November 2002 to October 2012 were used as test samples for the examination of model effects.

3.3 Model establishment

Through the use of pattern recognition and regression toolbox LIBSVM-3.23, the medium- and long-term scheduling model PSO-SVM of cascade reservoirs on the lower reaches of the Yalong River was established in MATLAB2018b. The model establishment steps can be summarized as follows:

(1) The model training samples and prediction samples were normalized. The normalization range is [0, 1]. The basic principle can be expressed as:

$$y = \frac{(y_{max} - y_{min}) \times (x - x_{min})}{(x_{max} - x_{min})} + y_{min} \quad (10)$$

Where x is the original sample data; y is the processed sample data; x_{max} , x_{min} , y_{max} , y_{min} are the maximum values of the corresponding processing data.

(2) The particle swarm algorithm parameters were set, and the training set samples were input so as to obtain the main parameters c and g of the multi-group LIBSVM learning model.

The main parameters of the particle swarm algorithm are: learning factors c_1 and c_2 , the general range of values [0, 4]; population size; maximum number of iterations; inertia weight. According to the empirical method, the particle swarm parameters of all PSO-SVM models are uniformly set as: learning factor $c_1=c_2=2$, and the maximum inertia weight is 0.9; the minimum inertia weight is 0.4; penalty coefficient upper bound ; relaxation coefficient $g \in (0.001, 100)$.

(3) The support vector machine and kernel function type in the training model of LIBSVM was selected to train the data training samples for the establishment of the learning model. Besides, the established learning model was used for the prediction of the training set, and the relative error and the qualified rate were analysed, so as to determine the optimal parameters c and g .

The training model parameters in LIBSVM are: support vector machine type, kernel function type, kernel function set value, and so on. In order to facilitate calculation and improve the prediction accuracy of the model, the selected support vector machine type in the training model parameters of PSO-SVM is multi-level classification and regression support vector machine, and the kernel function type is Gauss radial basis kernel

function, while all other parameters in the model were set to be default values.

(4) Regression prediction is performed on the prediction sample set according to the learning model established by the optimal parameters c and g .

4 Simulation Results and Analysis

4.1 Model Parameters

According to the above steps, the parameters of the PSO-SVM scheduling function model of each month period, of the cascade reservoirs in the lower reaches of the Yalong River are obtained, as shown in Table 1.

Table 1. Parameters of PSO-SVM scheduling model of cascade reservoirs in the lower reaches of the Yalong River

Project	Jinping Station I(JY)		Ertan Station(ET)	
Month	penalty coefficient $t c$	relaxation coefficient $t g$	penalty coefficient $t c$	relaxation coefficient $t g$
11	26.323	0.353	21.520	0.258
12	30.478	0.750	36.900	0.804
1	35.139	0.079	23.547	0.032
2	15.833	0.077	16.449	0.392
3	19.424	0.074	63.015	5.927
4	19.366	0.037	15.007	7.861
5	8.700	0.001	160.906	15.130
6	48.819	10.000	26.980	3.586
7	249.544	3.821	80.913	3.372
8	37.041	0.172	164.354	7.413
9	16.875	0.171	17.491	0.473
10	28.898	0.898	43.917	0.383

4.2 Comparative Analysis

In order to facilitate the comparison analysis concerning the water level prediction results of the linear regression and threshold regression models in literature [3], the mean relative error index was used to represent the degree of fitting with the optimal scheduling trajectory [7], and the PSO-SVM prediction yield was compared with linear regression model and the threshold regression model (the relative error less than 10% is considered qualified).

Taking April as an example, the PSO-SVM scheduling fitting results of the cascade reservoirs on the lower reaches of the Yalong River were calculated, as shown in Table 2. The results were compared with the linear regression and threshold regression models in the literature [3], as shown in Table 3. The absolute difference is the difference between the PSO-SVM model and the linear regression model or the threshold regression model.

Table 2. Water level prediction effects of PSO-SVM scheduling model for cascade reservoirs in the lower reaches of Yalong River in April (%)

Projects	PSO-SVM			
	Jingpin I (JY)		Ertan (ET)	
	Calibration period	Inspection period	Calibration period	Inspection period
Maximum error	12.13	13.33	5.24	10.66
Minimum error	0.02	0.00	0.00	0.46
Mean relative error	4.06	5.47	0.17	5.02
Qualification rate	92	90	100	90

Table 3. Comparison of two scheduling models for cascade reservoirs in the lower reaches of Yalong River in April (%)

Projects		Jinping Station I (JY)				Ertan (ET)			
		Calibration period	Absolute difference	Inspection period	Absolute difference	Calibration period	Absolute difference	Inspection period	Absolute difference
Linear regression	Maximum errors	12.33	-0.20	14.97	-1.64	10.89	-5.65	12.11	-1.45
	Minimum errors	0.43	-0.41	1.21	-1.21	0.01	-0.01	2.30	-1.84
	Mean relative errors	7.59	-3.53	8.36	-2.89	5.91	-5.74	7.50	-2.48
	Qualification rates	91.84	0.16	80.00	10.00	93.88	6.12	90.00	0.00
Threshold regression	Maximum errors	10.91	1.22	10.28	3.05	10.08	-4.84	11.70	-1.04
	Minimum errors	0.17	-0.15	0.38	-0.38	0.01	-0.01	1.64	-1.18
	Mean relative errors	6.41	-2.35	6.93	-1.46	4.64	-4.47	6.77	-1.75
	Qualification rates	93.88	-1.88	90.00	0.00	95.90	4.10	90.00	0.00

Table 3 shows that the maximum errors, minimum errors and average errors of PSO-SVM scheduling model are smaller than those of linear regression model by comparing PSO-SVM scheduling model with linear regression method, and the qualification rates are also higher than those of linear regression model. Compared with threshold regression, the average errors of PSO-SVM scheduling model are all less than those of threshold regression, and the qualification rates are generally better, so the scheduling fitting of PSO-SVM scheduling model degree is better; however, because the partial qualification rates of PSO-SVM scheduling model are less than those of the threshold regression model and the partial errors are greater than those of the threshold regression model, the fitting result of PSO-SVM scheduling model is more discrete and the stability of the simulation calculation is not as good as that of the threshold regression model. In summary, the effect of

PSO-SVM scheduling model based on support vector machine is better, which is equivalent to that of the threshold regression method.

4.3 Simulation operation

By using PSO-SVM scheduling model and C# language programming on VS2017 platform, the rolling simulation calculation of cascade hydropower stations in the lower reaches of Yalong River from November 2005 to October 2006 is carried out. The simulation results of Jinping I and Ertan were obtained. The relationship between water level change process, output and power generation of the two reservoirs and time is shown in Table 4 and Figure 1.

Table 4. Forecast water level, output and power generation from November 2005 to the end of October 2006

Projects	Jingpin I (JY)			Ertan (ET)		
Months	Water level (m)	Output (MN)	Power generation (MW·h)	Water level (m)	Output (MN)	Power generation (MW·h)
11	1876.9	2041	1469715	1200.0	1910	1375378
12	1870.5	1530	1138361	1200.0	1398	1040365
1	1838.9	2251	1674786	1200.0	2059	1532037
2	1813.3	1485	1033574	1196.8	1747	1215769
3	1800.0	859	639383	1167.8	2097	1560499
4	1800.0	630	453847	1161.2	820	590142
5	1816.1	608	451981	1176.6	154	114867
6	1843.1	2642	1902144	1180.8	2573	1852584
7	1850.3	3087	2297058	1188.5	2952	2195936
8	1863.3	1359	1011050	1191.7	1343	999135
9	1880.0	1798	1294659	1200.0	1526	1098691
10	1880.0	2378	1769335	1200.0	2639	1963470

Figure 1. Simulation operation results of the two reservoirs from early November 2005 to the end of October 2006

The results show that the water level of Jinping I Reservoir in the upstream reaches the dead water level evenly during dry season, while the water level of Ertan Reservoir in the downstream keeps relatively high at the beginning of the dry period and then reaches drawdown level centrally at the later stage, which is beneficial to improve the utilization efficiency of water resources. This is consistent with the research results of existing scheduling rules [11, 12]. In addition, the output in different time intervals of cascade power stations is quite different, but there is no phenomenon that the output of a certain period is too small. Therefore, the simulation results are reasonable so that the simulation operation can guide the actual operation of the power station.

5 Conclusions

This paper explores the cascade hydropower station scheduling model based on the technology of support vector machine (SVM), and PSO algorithm is used to

optimize the parameters of the model. Taking the lower reaches of Yalong River as an example, the results show that the simulation effect of cascade reservoir scheduling function based on PSO-SVM method is better than that of linear regression method but equivalent to that of nonlinear threshold regression technology. In addition, according to the result of actual simulation operation, the result that the scheduling model guides the operation of cascade hydropower stations meets the scheduling rule. Therefore, the PSO-SVM scheduling function model can provide reference for the actual operation of cascade hydropower stations.

Acknowledgements

We are very grateful for financial support from the National Key Research & Development Plan (2018YFB0905204), and Full-Time Postdoctoral Research & Development Fund of Sichuan University (2018SCU12062).

References

1. YOUNG G K. Finding reservoir operating rules [J]. Journal of the Hydraulics Division, 1967, 93(6): 297-322.
2. Changming Ji, Ting Zhou, Liping Wang, Yueyue Qin. Review of hidden long term stochastic optimal operation of reservoir hydropower stations [J]. Automation of Electric Power Systems. 2013, 37(16): 129-135.
3. Jidong Li, Weibin Huang, Guangwen Ma, Qingxu Zhao. Study on implicit stochastic joint optimal scheduling function of cascade reservoirs in the lower reaches of Yalong River [J]. International Journal Hydroelectric Energy. 2014, 32(12): 49-53.
4. Yongyi Chen, Qiufen Xiong. Application tutorial of support vector machine method [M]. China Meteorological Press. 2011.07, p6
5. Xianfeng He, Dongjian Zheng, Yanchang Gu. Threshold regression prediction model for dam safety monitoring and its application [J]. Journal of Yangtze River Scientific Research Institute. 2007(03): 20-22.
6. Cortes C, Vapnik V. Support-vector networks [J]. Machine Learning, 1995, 20(3):273-297.
7. Fangyuan Liu, Shuihua Wang, Yudong Zhang. Overview of support vector machine models and Applications [J]. Computer Systems and Applications. 2018, 27(04):1-9.
8. Jieying Yang. Particle swarm optimization and its application [M]. Beijing Institute of Technology press. 2017.03, p9.
9. Xiaoxun Zhu, Zhonghe Han. LS-SVM wind speed forecasting method based on PSO parameter optimization [J]. Journal of Chinese Electrical Engineering Science. 2016, 36(23): 6337-6342+6598.
10. Jinyin Chen, Hui Xiong, Haibin Zheng. Parameter optimization of support vector machine based on particle swarm optimization [J]. Computer Science. 2018, 45(06): 197-203.
11. Jidong Li, Weibin Huang, Qingxu Zhao, Guangwen Ma. Study on combined operating water level control mode of cascade reservoirs [J]. System Engineering Theory and Practice. 2016, 36(06): 1625-1632.
12. Wenhui Wu, Shuanghu Zhang, Zhongbo Zhang, Yunzhong Jiang, Rui Zhang. Evaluation of cascade reservoirs' centralized scheduling power generation efficiency -- Taking Wujiang cascade reservoirs as an example [J]. Journal of Hydroelectric Engineering. 2015, 34(10):60-69.