Study on Runoff Prediction of Representative Stations in Qiantang River Basin based on Particle Swarm Optimization and Support Vector Machine

Jie Sun [1], Jixiang Zhang [1,*], Shijian Zheng [2], Zhixia Li [3]

[1] Zhejiang Design Institute of Water Conservancy & Hydro Electric Power, Hangzhou 310002, China

[2] Taizhou water resources and Hydropower Survey and Design Institute Co., Ltd., Taizhou, 317000, China

[3] Linhai Water Resources Bureau, Taizhou, 317000, China

* Corresponding author e-mail: zdwp@zjwater.gov.cn

Abstract. Qiantang River is the largest river in Zhejiang Province and the mother river of Zhejiang. The middle and upper reaches of the basin are the Jinqu basin of Zhejiang Province. On both sides of the estuary are Hangjiahu Plain and xiaoshaoning plain. It is a gathering place of population and economic factors in Zhejiang Province, with a population of 10.8 million. Frequent floods in the Qiantang River Basin have caused serious losses to people's lives, property and social economy on both sides of the river. The amount of water from Qiantang River is related to the harm of flood and waterlogging natural disasters, but also affects the utilization of water resources and the future social and economic development of Qiantang River area. This paper attempts to combine the support vector machine model with the global optimization search method PSO model. PSO algorithm is used to optimize the penalty factor C and kernel parameters of support vector machine to improve the accuracy of the model. The calculation shows that the calculation accuracy of this model is much higher than that of artificial neural network model. This model can be used to predict the runoff of representative stations in Qiantang River Basin.

Keywords: Qiantang River Basin; PSO Algorithm; Flood Reduction.

1. Introduction

Qiantang River is the largest river in Zhejiang Province and the mother river of Zhejiang. The drainage area above Luchaogang sluice is 55491km2, and the main stream is 609km long, including 44467km2 in Zhejiang Province, also known as the mother river in Zhejiang. Zhejiang has 27 counties (cities and districts) in five districts, namely, Hangzhou, Quzhou, Jinhua, Shaoxing and Lishui. The middle and upper reaches of the basin are the Jinqu basin of Zhejiang Province. On both sides of the estuary are Hangjiahu Plain and xiaoshaoning plain. It is a gathering place of population and economic factors in Zhejiang Province, with a population of 10.8 million. Jinqu basin is the second commercial grain base in Zhejiang Province. At present, cotton production plays an important role in Zhejiang Province. In recent years, the socio-economic development in the region has been rapid. It is the region with the most potential for economic development in Zhejiang Province, especially the Hekou region. It is the south wing of the Yangtze River Delta Development Zone. It is surrounded by three rapidly developing cities of Shanghai, Hangzhou and Ningbo, with good development prospects. Frequent floods in the Qiantang River Basin have caused serious losses to people's lives, property and social economy on both sides of the river. From 1949 to 1999, there were 12 years of major disasters, with an average annual affected area of 732900 Mu and 34.7 deaths. The flood disaster years with peak discharge of more than 10000m3 / s at Lanxi station in the middle reaches of Qiantang River include 13 years, such as 1952, 1954, 1955, 1961, 1969, 1971, 1989, 1992, 1993, 1994, 1997, 1998 and 1999. After 2000, with the construction and implementation of flood control projects in the basin, the number of floods with great impact on the basin has decreased significantly, but some rainfall still has a great impact on the basin or local areas. Typical are the local Meiyu in 2008, the MeiXun flood in June 2011 and the MeiXun flood in June 2017.

The amount of water from Qiantang River is related to the damage degree of flood and waterlogging natural disasters, as well as the utilization of water resources [1], as well as the future social and economic development of Qiantang River area [2]. There are many runoff prediction methods at home and abroad [3] [4], but the accuracy of runoff prediction has been a problem perplexing many scholars. This paper is an attempt to combine the support vector machine model [5][6][7] with the global optimization search method PSO [8] model. PSO algorithm is used to optimize the penalty factor C and kernel parameters of support vector machine to improve the accuracy of the model.

2. Introduction to Support Vector Machine Model

Support vector machine was proposed by at & tbell laboratory research group led by vanpik in 1963. It is a pattern recognition method based on statistical theory. It has advantages in solving small sample, nonlinear and high-dimensional pattern recognition.

The input vector is mapped into the high-dimensional feature space f through a nonlinear mapping ϕ, and linear regression is performed in the feature space f, that is:

$$y = f(x) = (\omega \cdot x) + b \tag{1}$$

Where, $\omega \in R^n$ is the weight; Enter a value $x \in R^n$ for the sample; $b \in R$ Is the threshold. Parameters ω and b are determined by minimizing the structure:

$$\text{Min } \{\frac{1}{2}\|\omega\|^2 + C \cdot (v\varepsilon + \frac{1}{l}\sum_{i=1}^{l}(\varepsilon_i + \varepsilon_i^*))\} \tag{2}$$

$$\text{s.t} \quad ((\omega \cdot x_i) + b) - y_i \le \varepsilon + \varepsilon_i$$

$$y_i - ((\omega \cdot x_i) + b) \le \varepsilon + \varepsilon_i^* \tag{3}$$

$$\varepsilon_i^* \ge 0, \quad \varepsilon_i \ge 0$$

Where $\frac{1}{2}\|\omega\|^2$ is to increase the classification interval as much as possible to improve the generalization ability; $(v\varepsilon + \frac{1}{l}\sum_{i=1}^{l}(\varepsilon_i + \varepsilon_i^*))$ is the training error term. Where C is adjustable, and the greater C is, the heavier the penalty for error is. Therefore, a compromise can be made between the complexity of the algorithm and the error of the sample.

According to the strong duality theory, Lagrange multipliers α_i and α_i^* ($i = 1, 2, \cdots l$) are introduced to establish Lagrange functions, and partial derivatives of ω, b, ε_i and ε_i^* are obtained to make them equal to 0. Then the duality problem of the original problem is obtained as follows:

$$\min_{\alpha^{(*)} \in R^{2L}}$$

$$\frac{1}{2}\sum_{i,j=1}^{l}(\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j)K(x_i \cdot x_j) + \varepsilon\sum_{i=1}^{l}(\alpha_i^* + \alpha_i) - \sum_{i=1}^{l}y_i(\alpha_i^* - \alpha_i) \tag{4}$$

$$s \cdot t \quad \sum_{i=1}^{l}(\alpha_i^* - \alpha_i) = 0$$

$$0 \le \alpha_i^*, \alpha_i \le C, \ i = 1, 2, \cdots l \tag{5}$$

Where $K(x_i, x_j) = \phi(x_i)\phi(x_j)$ is the kernel function. Suppose the optimal solution $\overline{\alpha}^{(*)} = (\overline{\alpha}_1, \overline{\alpha}_1^*, \cdots, \overline{\alpha}_l, \overline{\alpha}_l^*)^T$ of the dual problem, obtain $\overline{b}$ according to the KKT condition, and construct the linear regression equation.

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

For the support vector machine model using radial basis kernel function, there are two main parameters that have a great impact on it, penalty factor C and radial basis kernel function parameter σ. The penalty factor C is mainly to adjust the proportion of trust range and empirical risk in the sample space, so as to make the generalization ability of support vector the best. The optimal value of C in different sample spaces is different. If the value of C is too small, it will be "under learning", otherwise it will be "over learning". The dimension of sample space is mainly changed by changing the mapping function. If the appropriate spatial dimension cannot be found, the generalization ability of support vector machine will cross. Therefore, it is very important to select the appropriate C and σ.

3. Runoff Prediction Model of Qiantang River Basin based on PSO Optimized Support Vector Machine

The algorithm flow of Qiantang River basin runoff prediction model based on PSO optimized support vector machine is as follows:

(1) Let the number of iterations $n=1$, initialize the particle swarm randomly, let the position vector of the i th particle be x_{id}^n, the velocity vector be v_{id}^n ($1\le i\le m, 1\le d\le D$), m is the population size of the particle swarm, and D is the dimension of the search space;

(2) Take the position vector x_{id}^n of each particle as the parameters c and σ of the support vector machine, calculate the estimated output f of the support vector machine, calculate the mean square error $RMSE$ at this time, and take it as the particle fitness value $RMSE_i^n$.

(3) Compare the current fitness value of each particle with its own best fitness value $RMSE_i^n$, if $RMSE_i^n < gbest$,so $gbest = RMSE_i^n$, $p_g^n = x_i^n$.

(4) Compare the current fitness value $RMSE_i^n$ of each particle with the best fitness value of the particle swarm $gbest$. If $RMSE_i^n < gbest$, then $gbest = RMSE_i^n$, $p_g^n = x_i^n$.

(5) Update the position vector x_{id}^{n+1} and velocity vector v_{id}^{n+1} of particles according to the update equation of particle swarm optimization algorithm; $n=n+1$, Returns (2) up to the maximum number of iterations.

4. Example Application

4.1 Qiantang River Basin Overview

Qiantang River has two sources: South and North. The main stream of Nanyuan Lanjiang River is 302.5km long, with a drainage area of 19468km2. It originates from the east foot of qingzhidaijian, Xiuning County, Anhui Province, and the elevation of the main peak is 1144m; It originates from Xin'an River in the north, with a main stream length of 358.9km and a drainage area of 11674km2. It originates from the north foot of Liugujian, Xiuning County, Anhui Province, and the elevation of the main peak is 1630m. The South and North sources converge in Meicheng, Jiande City. After merging, they are called Fuchun River, which flows to the east by North, passes through qililong Canyon, Fuchun River Power Station Dam, Tonglu and Fuyang to the mouth of yuanpudong River in Hangzhou, and joins Puyang river. It is called Qiantang River. It flows through Hangzhou, Haining Yanguan and Haiyan Shupu, and flows into Hangzhou Bay.

The drainage area above the gate of Qiantang River is 41945 km2. Including 35500 km2 in our province, 6200 km2 in Anhui Province, 109 km2 in Jiangxi Province and 136 km2 in Fujian Province. There are more than 130 tributaries with an area of more than 100km2 above the gate, including 60 from Lanjiang River in the South and 33 from Xin'an River in the north. The tributaries of Beiyuan from upstream to downstream from the left bank include Hengjiang River, Lianjiang River, mianxi River, Changyuan River, dazhouyuan River, Yunyuan port, Dongyuan port, qingpingyuan River, Lianhua River, Changning River, etc; The tributaries from the right bank include Xiaoyuan, Yiyuan, xinlingshui, chashui, Guixi, Jieyuan River, Guocun River, Wuqiang River, Fenglin port, shanghaiyuan, shouchang River, etc. The tributaries of Nanyuan from upstream to downstream from the left bank include cuntou River, Maji River, Hongqiao River, Fangcun River, datouyuan, tongshanyuan, Zhixi River, Tashi River, youbu River, Chixi River, Ganxi River, etc. Tributaries from the right bank include Hetian River, Zhongcun River, chihuai River, Longshan River, longrao River, Nanmen River, Jiangshan port, Wuxi River, luozhangyuan, Xiashan River, Lingshan port, Sheyang River, Xinfan River, Houda River, Jinhua River, Meixi River, Daxi River, etc. The tributaries of Fuchun River and Qiantang River from the left bank include XuXi River, Qingzhu port, Fenshui River, Luzhu River, Qingyunpu River, Xinqiao River, etc; Tributaries from the right bank include Sandu River, Luzi River, Dayuan River, Huyuan River, Dayuan River, Puyang River, etc.

Qiantang River Basin is located in the subtropical monsoon climate area, and the flood in the whole basin is mainly formed by large-area Meiyu. According to the records of the county chronicles in the region, during the 805 years from 1144 (the 14th year of Shaoxing in the Song Dynasty) to 1949, there were 58 floods, rainstorms and floods, mostly in the plum rain season from May to June of the Gregorian calendar. According to historical flood investigation and relevant literature records, 10 catastrophic floods have occurred in Qiantang River Basin since 1416, and the peak discharge of Lanxi is more than 15000m^3 / s. except that the flood in 1416 was caused by typhoon rain, the other 9 catastrophic floods were formed by Meiyu. The floods in the lower reaches of Qiantang River, such as Puyang River, are mainly caused by typhoon and rainstorm.

4.2 Runoff Characteristics

Watershed runoff is formed by precipitation. Precipitation is mainly divided into spring rain, plum rain and Taiwan rain. It is roughly bounded by Meicheng. The precipitation in the areas above Meicheng is dominated by Meiyu. The main rainy season occurs from March to June or from April to July every year. The precipitation for four consecutive months accounts for 55% ~ 60% of the annual precipitation, and the maximum monthly precipitation occurs in May or June; The areas below Meicheng, including Fenshui River, Puyang River and Fuchun River, are obviously affected by tropical storms and typhoons. There are two rainy seasons in normal years. The first rainy season is from April to July, which is spring rain and plum rain. The precipitation in four months accounts for 40% ~ 45% of the annual precipitation, and the maximum monthly precipitation mostly occurs in June; The second rainy season is in August or September, which lasts a short time. It is caused by typhoons and tropical storms. The maximum monthly precipitation accounts for about 15% of the whole year. In years without typhoon (tropical storm) transit or peripheral impact, there is only one rainy season.

Controlled by precipitation, runoff is unevenly distributed throughout the year. The annual maximum runoff for four consecutive months can account for 50% to 60% of the annual runoff. Areas affected by typhoons and tropical storms account for a small proportion, but the unevenness of runoff distribution is more obvious. The interannual variation of runoff is also great. The runoff in dry years is only 50% ~ 60% of the multi-year average value, while in wet years it can reach 1.5 ~ 2 times of the multi-year average value.

4.3 Runoff Analysis

The monthly runoff series of Fuchunjiang station from 2005 to 2019 is selected as the research object. Take the data from 2005 to 2018 as simulation data to predict the data in 2019. Let the population size m of the particle swarm be 100, the inertia weight change linearly w from 1 to 0.5, set $V_{\max}$ to 0.5, the acceleration constant sum is set c_1 and c_2 to 2, and the number of iterations is 100.

Compare the simulation prediction results with the original sequence, and the comparison results are shown in the figure below:

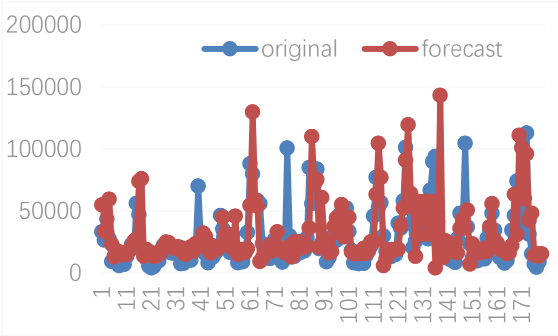

Figure 1. Comparison between simulation prediction results and original data

Calculate the relative error of the model according to the following formula:

$$\left|(x_i - x_{i0}) / x_{i0}\right| \quad i = 1, 2, \cdots, n \tag{7}$$

Where x_i is the simulated predicted value, the original value is x_{i0} and the number of sequences is n. Calculate the percentage of relative error less than a certain value according to the following formula.

$$\mathrm{n_i}/\mathrm{n_{all}} \tag{8}$$

$\mathrm{n_i}$ is the total number of samples whose relative error is less than a certain percentage, and $\mathrm{n_{all}}$ is the number of all samples in the statistical stage.

The evaluation results of particle swarm optimization support vector machine model are compared with those of neural network model. The comparison results are shown in the table below:

Table 1 Comparison of evaluation results of two models in simulation stage

Table 1. Comparison of evaluation results of two models in simulation stage

	Simulation phase		Prediction stage	
	Percentage of relative error < 10%	ercentage of relative error < 20%	Percentage of relative error < 10%	ercentage of relative error < 20%
This model	21.05%	83.04%	36.11%	86.13%
ANN model	16.81%	54.32%	32.35%	60.42%

5. Conclusion

In order to improve the runoff prediction accuracy of representative stations in Qiantang River Basin, PSO optimization support vector machine model is introduced in this paper. By comparing the calculation results of the model with the artificial neural network model, it shows that the model improves the accuracy of simulation and prediction stage, and the model can be used in the runoff prediction model of representative stations in Qiantang River Basin. High precision runoff prediction can provide technical support for flood control, disaster reduction and water resources utilization in Qiantang River Basin.

Acknowledgments

This work was financially supported by Water conservancy science and technology plan project of Zhejiang Province (Project No.RC2008) fund.

References

[1] Yang Daohui, Ma Guangwen, Liu Qifang, Tao Chunhua, Guo Xiaming. Application of BP network model based on particle swarm optimization algorithm in runoff prediction [J]. Journal of hydropower, 2006 (02): 65-68.

[2] Li Lin, Shen Hongyan, Dai Sheng, Xiao Jianshe, Shi Xinghe. Response of runoff in the source area of the Yellow River to climate change and prediction of future trend [J]. Journal of geography, 2011, 66 (09): 1261-1269.

[3] Feng Ping, Ding Zhihong, Han Ruiguang, Zhang Jianwei. Rainfall runoff neural network prediction model based on EMD [J]. System engineering theory and practice, 2009, 29 (01): 152-158.

[4] Zhang Shaowen. Analysis and prediction of natural annual runoff variation characteristics in the Yellow River Basin [D]. Sichuan University, 2005.

[5] Liao Jie, Wang Wensheng, Li Yueqing, Huang Weijun. Support vector machine and its application in runoff prediction [J]. Journal of Sichuan University (Engineering Science Edition), 2006 (06): 24-28.

[6] Yu Guorong, Xia Ziqiang. Chaotic time series support vector machine model and its application in runoff prediction [J]. Progress in water science, 2008 (01): 116-122.

[7] Tang Qi, Wang Hongrui, Xu Xinyi, Wang Cheng. Hydrological time series model based on mixed kernel function SVM and its application [J]. System engineering theory and practice, 2014, 34 (02): 521-529.

[8] Ding genhong. Research and application of intelligent optimization algorithm for reservoir flood control operation [D]. Hehai University, 2008.