Runoff simulation Based on Least Square Support Vector Machine

LIU Jun Ping , ZHOU Jun Jie & ZOU Xian Bai

College of Civil Engineering and Architecture, Zhejiang University of Technology,
Hangzhou ,China

KEYWORD: Statistical Learning Theory; Least Square Support Vector Machine; Runoff; Simulation

ABSTRACT: River runoff is highly nonlinear as affected by the combination of climate, underlying surface condition, etc. Prediction of runoff may guide engineering design, construction and operation. Statistical Learning Theory (SLT) studies the rules of machine learning with finite samples. Support Vector Machine (SVM) is a new machine learning method based on Statistical Learning Theory. It is a solution for the highly nonlinear classification and regression in sample space. Map the one-dimensional runoff series input space of one hydrologic station of the Yellow River onto high-dimensional input space. Then calculate the embedding dimension of runoff time series and reconstruct runoff series into three-dimensional phase space. Using radial base kernel function to learn from 83 training samples through grid search method and optimize model parameters to establish the Least Square Support Vector Machine (LSSVM) prediction model of river runoff. Fitting mean-square error of the model is 0.0148. Prediction mean-square error of the model on 20 samples is 0.0120, a correlation coefficient of 0.975 between predicted and measured values. The result shows that the generalization ability of LSSVM model is high and the prediction result is satisfactory..

INTRODUCTION

As affected by the combination of climate and underlying surface, such as landform, soil, vegetation and human activities, river runoff alternates between high-water period and low-water period. River runoff is mostly affected by climate. Underlying surface controls the transformation from precipitation to runoff by influencing evapotranspiration, infiltration and converging, thus governing the redistribution process of precipitation. Underlying surface factors such as vegetation and soil are closely related to climate as well. Therefore, variation in river runoff is bound up to global climate change. Global warming and sunspot activity in recent years have greatly impacted the pattern of river runoff. As economies are growing, human activities (e.g. large reservoir storage, trans-basin diversion, water-soil conservation and other development, utilization, treatment and configuration initiatives), various social & economic and environmental elements also have a growing influence on river runoff. River runoff shows quasi-periodicity and nonlinearity. With respect to nonlinearity, Neural Network, Chaos Theory, etc. are widely used. However, since Neural Network algorithm uses Empirical Risk Minimization Principle and it is prone to local optimum, the field application of Neural Network is limited. Conventional Time Series Prediction and Neural Network are both based on statistics. Conventional statistics studies the Asymptotic Theory when the number samples tends to infinity. Since in reality the number of samples is limited, some learning and prediction methods excellent in theory are not ideal in field application.

Statistical Learning Theory (SLT) is a theory that studies the rules of machine learning with finite samples. Vapnik (Vapnik 1995)and other scholars started the study in 1960s to 1970s. The theory has developed and matured through mid-1990s. On the other hand, other learning methods, such as Neural Network, are lack of substantial progress. Therefore, SLT has become an increasingly appreciated theory (Nello & John 2000,Shevade et al. 2000).

LEAST SQUARE SUPPORT VECTOR MACHINE

Support Vector Machine (SVM) is a new machine learning method proposed by V. Vapnik and his team of AT&T's Bell Labs based on SLT. SLT is a theory that studies the rules of machine learning with finite samples. SLT not only considers requirements on generalization ability, but also pursues the optimal result with limited information. It is a new theory in the field of machine learning and is currently the best theory for small sample statistical estimation and prediction learning. SLT studies the prerequisites of Empirical Risk Minimization Principle, the relationship between empirical risk and expected risk of limited sample, and how to use these theories to find out new learning principles and methods in a systematic way(Deng & TIAN 2004). SLT solves the issues of model selection and overfitting, nonlinearity and curse of dimensionality, local minimum, etc. to a great extent, establishing a favorable theoretical framework for small sample machine learning issues. Currently SVM has been successfully used in a number of research applications.

SVM prediction, i.e. Support Vector Regression (SVR), is based on the concept of mapping data onto high-dimensional feature space through nonlinear mapping and proceeding with linear regression in this space.

Least Square Support Vector Machine (LSSVM) is an algorithm developed from Standard SVM (S-SVM). Loss function is defined as error square sum, which is obtained by converting inequality constraints to equality constraints in S-SVM algorithm. LSSVM reduces computational complexity and improves convergence rate by solving a group of linear equations instead of the quadratic programming in S-SVM. LSSVM only needs to determine the shape parameter and penalty parameter without selecting the value of insensitive loss function. Therefore, LSSVM is easier to use than S-SVM(GUO et al.2006, LIN &CHENG 2006,LI et al.2007).

The fundamental of SVR is mapping input space onto high-dimensional feature space through nonlinear mapping and proceeding with linear regression in the high-dimensional space.

For nonlinear regression, establish training sample as $(x_i, y_i), \mathbf{L}, (x_n, y_n) \in R^n \times R, i = 1,2,...,l$, and the nonlinear regression function is:

$$y_i = w^T \mathbf{j}(x_i) + b + e_i \qquad (1)$$

S-SVM uses e insensitive loss function and LSSVM defines loss function as quadratic item with an error of e_i.

According to Structural Risk Minimization Principle, LSSVM risk function is:

$$J_l = \frac{1}{2} \|W\|^2 + g \sum_{i=1}^{l} e_i^2 \qquad (2)$$

Then the original optimization issue can be written as:

$$Min \quad \frac{1}{2}\|w\|^2 + \frac{1}{2} g \sum_{i=1}^{l} e_i^2 \qquad (3)$$

Constraint being:

$$y_i = w^T \mathbf{j}(x_i) + b + e_i \quad i = 1,2,\mathbf{L},l \qquad (4)$$

Where $\mathbf{j}(\cdot): R^n \to R^{nh}$ is the nonlinear function that maps data from original space to high-dimensional Hilbert feature space; w is weight vector; $e_i \geq 0$ as error variable, equivalent to the slack variable under S-SVM; $g > 0$ as adjustment parameter factor. g is a constant to control the penalty on out-of-error samples. It creates a medium solution between error and complexity, giving better generalization ability to the result function. The greater the value of g, the smaller the regression error of the model.

To solve the abovementioned optimization issue, Lagrange function can be introduced to transfer constrained optimization to non-constrained optimization.

$$L = \frac{1}{2}\|w\|^2 + \frac{1}{2}g\sum_{i=1}^{l}e_i^2 - \sum_{i=1}^{l}a_i(w^T j(x_i) + b + e_i - y_i) \qquad (5)$$

where a_i is the multiplier of Lagrange. Optimal a_i and b can be obtained through condition KKT (Karush-Kuhn-Tucker):

$$\left. \begin{aligned} \frac{\partial L}{\partial w} = 0 &\rightarrow w = \sum_{i=1}^{l}a_i j(x_i) \\ \frac{\partial L}{\partial b} = 0 &\rightarrow \sum_{i=1}^{l}a_i = 0 \\ \frac{\partial L}{\partial e_i} = 0 &\rightarrow a_i = ge_i \\ \frac{\partial L}{\partial a_i} = 0 &\rightarrow w^T j(x_i) + b + e_i - y_i = 0 \\ i = 1,2,\mathbf{L},l \end{aligned} \right\} \qquad (6)$$

From equation (3), the optimization becomes solving a linear equation as shown below:

$$\begin{bmatrix} 0 & 1 & 1 & \mathbf{L} & 1 \\ 1 & K(X_1,X_1)+1/g & K(X_1,X_2) & \mathbf{L} & K(X_1,X_n) \\ 1 & K(X_2,X_1) & K(X_2,X_2)+1/g & \mathbf{L} & K(X_2,X_n) \\ \mathbf{M} & \mathbf{M} & \mathbf{M} & \mathbf{O} & \mathbf{M} \\ 1 & K(X_n,X_1) & K(X_n,X_2) & \mathbf{L} & K(X_n,X_n)+1/g \end{bmatrix} \begin{bmatrix} b \\ a_1 \\ a_2 \\ \mathbf{M} \\ a_l \end{bmatrix} = \begin{bmatrix} 0 \\ y_1 \\ y_2 \\ \mathbf{M} \\ y_l \end{bmatrix} \qquad (7)$$

Where: $K(x_i,x_j)$ is the kernel function.

Solve equation (7) to obtain a and b, and the solution of the nonlinear regression function is:

$$f(x) = w^T f(x) + b = \sum_{i=1}^{l}a_i K(x_i,x) + b \qquad (8)$$

SELECTION OF OPTIMAL PARAMETERS FOR LSSVM MODELING

When modeling, first select the kernel function. Use linear, polynomial, radial base and Sigmoid kernel functions respectively to create the model. Comparison shows that the radial base kernel function (RBF) is suitable for the research topic of this article. If RBF is used, two parameters, g and s, need to be determined when solving the model. These two parameters are determined by Cross-Validation (CV)(GUO et al.2006).

Cross-Validation

CV includes k-fold Cross-Validation (k-CV) and Leave-one-out Cross-Validation (Loo-CV). k-CV is used in this article. k-CV procedures are listed below:

(1) Set cycles $t=1$;

(2) Randomly rearrange original input/output matrix;

(3) Divide training samples $(x_i, y_i), i = 1,2,\mathbf{L},n$ into k parts, approximately equal in numbers and do not intersect with each other, i.e. k-fold S_1, S_2, …, S_k.

Train and test the data for k times, i.e. iteration for k times. For the ith iteration, select S_i as test set and the union of the rest S_1, …S_{i-1}, S_{i+1}, …S_k as training set. Use the training set for model training

to obtain regression model. Then use the regression model to test the test set S_i. Error between the actual value and predicted value of S_i is represented by mean-square error:

$$\cos t_i = mse(f(x_j) - y_j), j \in S_i \qquad (9)$$

(4) After calculation of k times, there will be k errors of $\cos t_1, \cos t_2, \mathbf{L}, \cos t_k$. The average error of these k errors shall be:

$$error(t) = \frac{\sum_{i=1}^{k} \cos t(i)}{k} \qquad (10)$$

(5) Repeat step 2 through 4 for T times to obtain $error(1), error(2), \mathbf{L}, error(T)$. The average error of T cycles shall be:

$$error = \frac{\sum_{t=1}^{T} error(t)}{T} \qquad (11)$$

Equation (11) is the estimation of modeling error, i.e. k-fold CV error.

Grid Search Method

For parameters to be optimized, apply grid in the available given interval and calculate objective function value on the grid. Select the corresponding parameter of objective function that satisfies optimal property of modeling function as optimal parameter. For the LS-SVM model in this article, give the initial value for g and s. Then divide the range corresponding to these two parameters into 10 equal parts, i.e. a 10×10 grid. Calculate the value of $error$ on each grid node with CV method. Use the value of $error$ as the objective function of the corresponding node. Draw a contour line for each $error$ value of the 10×10 grid and solve g and s of the satisfactory contour line as optimal parameters.

LSSVM RUNOFF TIME SERIES PREDICTION

Model Input/Output

Prediction of time series assumes that the future time series can be predicted based on past time series. Runoff time series prediction model is to find out the nonlinear function between past time series and future time series. Use the monthly runoff data of one hydrologic station of the Yellow River main stream and reconstruction theory to reconstruct chaotic time series into a low order nonlinear dynamics system. Determine the embedding dimension of chaotic time series to reconstruct phase space. The reconstruction theory of state-space expands the vector of input space onto high-dimensional space, so as to extract the information and rule residing within the system. Calculated embedding dimension is 3 and there are 103 samples in the reconstructed state-space, as shown below:

$$\begin{bmatrix} y_1 \\ y_2 \\ \mathbf{M} \\ y_l \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & x_3 \\ x_2 & x_3 & x_4 \\ & \mathbf{L} & \\ x_l & x_{l+1} & x_{l+2} \end{bmatrix} \qquad (12)$$

Reconstruct the multi-dimensional state-space of runoff chaotic series to for learning sample series. Then utilize SLT's LSSVM to construct prediction model for the forecast of runoff. Use monthly runoff of the first three months, x_i, x_{i+1}, x_{i+2}, to forecast the runoff of the fourth month y_i, where $i = 1, 2, ..., l, l = 103$, i.e. SVM prediction model has three input variables and one output variable.

Select the first 83 samples for SVM model training and the last 20 samples as test samples to forecast the monthly runoff of these 20 samples. Before running the model, normalize the input and output of samples to contain the data within the range of [0, 1].

Runoff Time Series Prediction

With T=10 and k=10, middle calculating process is shown in Figure 1(Suykens et al.2002).

Through 10-fold VC of 10 cycles, the optimal parameters of LSSVM model are g =951.8036 and s^2 =0.0101. Learning of 83 samples shows that the fitting result is favorable, with a mean-square error of 0.0148. See Fig. 2 for fitting result. As shown in the figure, fitting value coincides with actual value well.
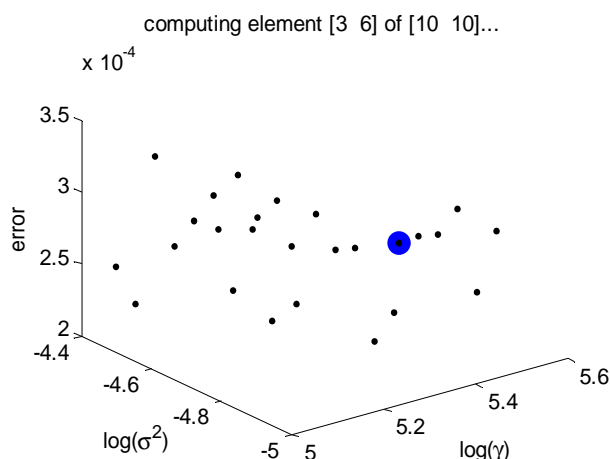

Figure 1. The result of grid search

Mean-square error of prediction samples is 0.0120. Minimum relative error is 0.57% and maximum relative error is 19.06%, average value being 8.97%. In mathematics, inner product can be used to indicate the similarity between two vectors and kernel function is used to describe the similarity between two vectors in feature space. Therefore, SVR can be seen as an algorithm for similarity issues. The correlation coefficient between the predicted and measure values of the 20 prediction samples is 0.975, indicating strong relevance.

Therefore, LSSVM model has good performance in river runoff predication, and its generalization ability is high. Computing speed of LSSVM model is fast and it shows favorable results in both fitting and predicting. LSSVM is a feasible prediction method.

Forecast the 20 prediction samples with trained LSSVM model. Forecast results are shown in Table 1. The results show that prediction value coincides with actual value well.

SUMMARY

To forecast runoff, a 3-input & 1-output SVM predication model is established. SVM algorithm realizes structural risk minimization. It seeks for a medium solution between model complexity and learning ability based on limited samples, ensuring better generalization ability. The method features high predication accuracy and generalization ability with small training samples. The predication result shows that the runoff time series prediction model of LSSVM has high predication accuracy, providing a new method for runoff forecast.

Table 1. The predicting error of SVM model

Sample	Actual runoff m³/s	Predicting runoff m³/s	Relative error %
1	144	164	14.00
2	211	216	2.30
3	412	334	18.91
4	560	520	7.21
5	498	491	1.36
6	236	270	14.29
7	171	141	17.43
8	185	171	7.38
9	224	259	15.79
10	365	360	1.40
11	597	594	0.57
12	864	984	13.87
13	772	625	19.06
14	681	690	1.38
15	597	577	3.29
16	179	201	12.41
17	134	139	3.78
18	136	142	4.38
19	212	235	10.87
20	311	341	9.61

REFERENCES

1) Deng Nai-yang,TIAN Ying-jie.New Method in Data Mining—Support Vector Machine[M].Beijing:Science press(in Chinese),2004:125-160.

2) GUO Hui,LIU Heping,WANG Ling.Method for Selecting Parameters of Least Squares Support Vector Machines and Application[J].Journal of System Simulation(in Chinese), 2006，18(7)：2033-2036.

3) LI Xiaodong,XI Shengyang,PAN Li.Forecast of China Grain Production Based on Least Squares Support Vector Machine[J]Research of Soil and Water Conservation(in Chinese), 2007，14(6)：329-331.

4) LIN Jianyi CHENG Chuntian. Application of support vector machine method to long-term runoff forecast[J]. Journal of Hydraulic Engineering(in Chinese),2006，37(6)：681-686.

5) Nello Cristianini,John Shawe-Taylor.An Introduction to Support Vector Machines and other Kernel-based Learning Methods[M].London:Cambridge University Press,2000.

6) Shevade S K,Keerthi S C.Bhattacharyy et al.Improvements to SMO algorithm for SVM regression[J].IEEE Trans on Naural Networks,2000,11(5):1188-1193.

7) Suykens J.A.K., Van Gestel T., De Brabanter J.,et al. Least Squares Support Vector Machines[M]. Singapore:World Scientific, 2002

8) Vapnik V. N.. The Nature of Statistical Learning Theory[M]. New York: Springer, 1995:70-256.