

Papers published in *Hydrology and Earth System Sciences Discussions* are under open-access review for the journal *Hydrology and Earth System Sciences*

**The support vector
machines approach**

H. Zhou et al.

Ice breakup forecast in the reach of the Yellow River: the support vector machines approach

H. Zhou, W. Li, C. Zhang, and J. Liu

School of Civil and Hydraulic Engineering, Dalian University of Technology,
Dalian 116024, China

Received: 21 March 2009 – Accepted: 23 March 2009 – Published: 9 April 2009

Correspondence to: W. Li (dlutliwei@gmail.com)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Abstract

Accurate lead-time forecast of ice breakup is one of the key aspects for ice flood prevention and reducing losses. In this paper, a new data-driven model based on the Statistical Learning Theory was employed for ice breakup prediction. The model, known as Support Vector Machine (SVM), follows the principle that aims at minimizing the structural risk rather than the empirical risk. In order to estimate the appropriate parameters of the SVM, Multiobjective Shuffled Complex Evolution Metropolis (MOSCEM-UA) algorithm is performed through exponential transformation. A case study was conducted in the reach of the Yellow River. Results from the proposed model showed a promising performance compared with that from artificial neural network, so the model can be considered as an alternative and practical tool for ice breakup forecast.

1 Introduction

Ice flood is a common phenomenon in river of north china every year. It often occurs during the period of thaw when river discharge increases due to snowmelt that causes the forces on an ice cover exceed its strength. Ice jams are often formed when the broken ice is transported along the river and accumulated in the river bed. It can cause many hazards, such as bridge or levee failure, structural damage of dams, and erosion of riverbed and banks (Massie et al., 2002). Ice breakup prediction can be used to increase warning time and to minimize damage caused by ice floods or ice jams. Therefore, it is of great importance to make an accurate ice breakup forecast.

Ice breakup prediction is much difficult because the formation of breakup result from a complex interaction between hydrologic, hydraulic, and meteorological processes. Due to the complexity of physical processes that cannot currently be described with deterministic models, there are no reliable methods to predict the date of breakup with a significant lead time (White, 2003). Many models have been developed for

HESSD

6, 3175–3198, 2009

The support vector machines approach

H. Zhou et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



predicting the ice breakup, including empirical methods, mathematical models, and statistical models et al.

White provided examples of existing breakup ice jam prediction methods and discussed their potential advantages and disadvantages (White, 2003). Most advances in river ice hydrology were reviewed in other literatures (Beltaos, 2000; Morse and Hicks, 2005). Beltaos calculated the threshold flows that can result in significant jam flooding using a numerical model named RIVJAM (Beltaos, 2003). For statistical models, logistic regression was widely used for predicting breakup, ice jam occurrence (White, 1996). Massie pointed out that breakup ice jam prediction models had historically been limited to classical empirical single-variable threshold-type analyses to statistical methods. He employed neural network for ice jam prediction and improved the forecast accuracy (Massie et al., 2002). Instead of using neural network alone, fuzzy logic and artificial neural networks were applied for modeling the maximum water level during river ice breakup for both flood and non-flood event years (Mahabir et al., 2006). A fuzzy optimization neural network approach was also developed for forecasting freeze-up date and break-up date (Chen and Ji, 2005). In China, the index method was firstly used for ice breakup forecast in the 1950', it is a empirical method which has the disadvantages of short lead time and poor accuracy. To solve problems aforementioned, empirical correlation method is developed based on certain physical processes, it improved the forecast accuracy but usually was restricted to the application conditions. In the 1990', some mathematical models were developed which is effective and efficient, but it is often failed due to some parameters can not be achieved in practice which limited their wide applications.

In summary, although physically based models have several advantages, but they are so complex and require many hypothesis for establishment, so black box model have gained much popularity in recent years. It doesn't need to understand the complex process and the only thing is to establish the relationship between input and output variables which make it ease of use. In the past decade, artificial neural network (ANN), as one of the famous black-box models, has been increasingly applied to vari-

The support vector machines approach

H. Zhou et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



ous hydrological problems for it can dealing with complex nonlinear processes (ASCE, 2000b).

But there are some limitations for the widely application of ANN, such as the network architecture selecting, training algorithm selecting, and the most serious problems is that it do not lead to one global or unique solution (ASCE, 2000a).

In this paper, an alternative data-driven model based on statistical learning theory called Support Vector Machine (SVM) is used for the task of ice breakup prediction. It has originally been used for classification purpose and later was been extended for regression and prediction (Huang et al., 2005; Ye et al., 2005; Niu et al., 2006; Sun and Yang, 2006). SVM follows the principle of structural risk minimization. The key property of SVM is that it prevents the overfitting (or overtraining) and the solution is always unique and globally optimal. As for the optimal parameters selecting of SVM, the Multiobjective Shuffled Complex Evolution Metropolis (MOSCEM) algorithm is employed.

The remainder of the paper is organized as follows: Sect. 2 provides a briefly introduction to the SVM and its parameters identification method (MOSCEM). A description of study area and data used is presented in Sect. 3. In Sect. 4, the implementation of SVM for ice breakup prediction is described and results are shown compared with ANN. The paper closes with a summary and some problems to be studied.

2 Support vector machine for regression (SVR)

2.1 Introduction to SVR

Here, a brief description of SVM for regression is given below. Detailed descriptions of SVR can be found in Vapnik (1995).

The SVR maps the input data x into a high-dimensional feature space F by linear or nonlinear mapping, in which the training data may exhibit linearity, and then to perform linear regression in this feature space. The regression estimation is to construct

The support vector machines approach

H. Zhou et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



The support vector machines approach

H. Zhou et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



a function according to a given data set $\{(x_i, y_i)\}_n$, where x_i denotes the input vector; y_i denotes the output value and n is the total number of data sets. The aim is to identify a regression function $y=f(x)$, which can accurately predict the outputs $\{y_i\}$ corresponding to a new set of input–output examples, $\{(x_i, y_i)\}$. The linear regression function using the following function:

$$f(x) = \omega \cdot \phi(x) + b, \quad (1)$$

where $\phi(\cdot)$ is a nonlinear function by which x is mapped into a feature space, b denotes the bias. Linear regression is performed in the high-dimensional feature space by ε -insensitive loss function. To prevent over-fitting, in other words, improve the generalization ability, objective function (2) is established, which is minimizing summation of the empirical risk and a complexity term $\|\omega\|^2/2$. The coefficients ω and b can thus be estimated by function (2).

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

$$L_\varepsilon(f(x), y) = \begin{cases} |f(x) - y| - \varepsilon & |f(x) - y| \geq \varepsilon \\ 0 & \text{otherwise} \end{cases}, \quad (3)$$

where $\|\omega\|^2/2$ is the regularization term which denotes the Euclidean norm, and $L_\varepsilon(\cdot)$ is called ε -insensitive loss function that measuring the empirical risk, C is a positive constant that determines the trade-off between the model complexity and the amount up to which error larger than ε are tolerated.

The points lying on or outside the ε -bound of the decision function are support vectors (black points). On the right, the ε -insensitive loss function is shown in which the slope is determined by C .

To estimate ω and b , Eq. (2) is converted to the primary function given by Eq. (4),

$$\begin{aligned} \min R &= \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \\ \text{subjected to} &\begin{cases} y_i - \omega \cdot \phi(x) - b \leq \varepsilon + \xi_i \\ \omega \cdot \phi(x) + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases}, \end{aligned} \quad (4)$$

where ξ_i and ξ_i^* are slack variables that specify the upper and the lower training errors subject to an error tolerance ε . In this optimization problem, most data examples are expected to be in the ε -tube as show in Fig. 1. If a data example $\{(x_i, y_i)\}$ is outside the tube, then an error ξ_i or ξ_i^* exists. From Eq. (4), some characters of SVR can be found: (i) the training error is minimized by minimizing ξ_i and ξ_i^* and (ii) $\|\omega\|^2/2$ is minimized to raise the flatness of $f(x)$.

Equation (4) is a standard optimal problem which can be solved by applying Lagrange theory. By introducing Lagrange multipliers α_i, α_i^* , Eq. (4) can be converted to the following form:

$$\max R(\alpha_i, \alpha_i^*) = -\frac{1}{2} \sum_{i,j=1}^n (\alpha_i - \alpha_i^*)(\alpha_j^* - \alpha_j)(\phi(x_i), \phi(x_j)) + \sum_{i=1}^n \alpha_i^*(y_i - \varepsilon) - \sum_{i=1}^n \alpha_i(y_i + \varepsilon). \quad (5a)$$

Subject to constraints

$$\sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0, \quad 0 \leq \alpha_i, \quad \alpha_i^* \leq C. \quad (5b)$$

To solve the optimal problem above, the Lagrange multipliers α_i, α_i^* are calculated, and the optimal desired weight vector of the regression hyperplane is

$$\mathbf{w} = \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(\phi(x), \phi(x_i)). \quad (6)$$

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Thus, the regression function can be written as,

$$f(x, \alpha, \alpha^*) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(\phi(x), \phi(x_i)) + b. \quad (7)$$

Herein, $K(\phi(x), \phi(x_i))$ is called the Kernel function. The value of the Kernel is the inner product of the two vectors x_i and x_j in the feature space $\phi(x_i)$ and $\phi(x_j)$. Any function that satisfies Mercer's condition (Vapnik, 1998) can be used as the Kernel function.

By using different kernel functions listed in Table 1, the SVR algorithm can construct a variety of learning machines.

In Eq. (5), only some of $(\alpha_i - \alpha_i^*)$ will be held as non-zero values, they are called the support vector. That is, these data points lie on or outside the ε -tube (Fig. 1).

The parameters of SVR are the cost constant C , the radius of the insensitive tube ε , and the kernel parameters. These parameters are mutually dependent. Parameter C controls the smoothness of the approximation function. A greater C value indicates that the objective is mainly to minimize the empirical risk, which makes the learning machine complex. On the contrary, a smaller C value makes learning machine yield poor approximation. The parameter ε also affects the smoothness or complexity of the approximation function. In addition, ε determines the number of support vectors because it controls the accuracy of the approximation function. Smaller values of ε may lead to more support vectors and result in a complex learning machine and vice visa. So it is very critical to set the appropriate SVR parameters.

2.2 SVR parameters selection

Many algorithms have been recorded in literature for SVM parameters selection, such as genetic algorithm (GA), particle swarm algorithm (PSO) and Shuffled Complex Evolution algorithm (SCE-UA) (Duan, 1992). Single objective functions are widely used for SVM training, however, practical experience suggests that single-objective functions,

The support vector machines approach

H. Zhou et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



The support vector machines approach

H. Zhou et al.

no matter how carefully chosen, are often insufficient to measure all the characteristics of the training error (Vrugt et al., 2003). In this study, an effective and efficient algorithm for multiobjective optimization, entitled the Multiobjective Shuffled Complex Evolution Metropolis (MOSCEM) algorithm developed by Vrugt (2003), was employed for SVM parameters selection. MOSCEM is an improvement over the Shuffled Complex Evolution (SCE-UA) algorithm, using Pareto dominance concept to evolve the population toward Pareto optimal solutions and merging the strengths of Metropolis algorithm.

In this section, for optimizing the following multiobjective function, the implementation framework of MOSCEM algorithm (Vrugt, 2003) is briefly introduced below.

$$\min_{\theta} F(\theta) = \{F_1(\theta), F_2(\theta), \dots, F_n(\theta)\}, \quad (8)$$

in which $F_i(\theta)$ is a single objective function, θ is the parameter set to be optimal selected.

Steps:

- (1) Generate sample: Generate s samples $\{\theta_1, \theta_2, \dots, \theta_s\}$ randomly from the feasible parameter space and compute the multiobjective vector $F(\theta_i)$ at each point θ_i .
- (2) Ranking points: Compute the fitness f_i for each individual of the sample using the Pareto concept, sort the s points in order of decreasing fitness value and store them in an array $D[1:s, 1:n+M+1]$, where n is the number of parameters, and the remaining $M+1$ columns are used to store the multiobjective vector and the fitness values.
- (3) Complexes generating: Partition D into q complexes C^1, C^2, \dots, C^q , each containing m points.
- (4) Sequence evolution: A new candidate point in each sequence k is generated using a multivariate normal distribution centered around the current draw of sequence (k) augmented with the covariance structure induced between the points in complex k . Metropolis acceptance rule is used to decide whether the offspring

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



The support vector machines approach

H. Zhou et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



should be added to the current sequence or not. If it is accepted, then the worst member of the current complex k will be replaced with it, otherwise, the worst member should be replaced with the last member of S^k . Finally, when predefined number of iterations reached, new complexes are formed by means of shuffling process.

- (5) Unpack all complexes C back into D and calculating the fitness of all points, sort the points in order of increasing fitness value.
- (6) Check the convergence criteria. Once convergence criteria are satisfied, stop; Otherwise, go to step 3.

MOSCEM algorithm guarantees convergence toward the Pareto set of solutions which reflect the model structure uncertainty, so the probability forecasts can be made rather than point forecasts by traditional method. In MOSCEM algorithm, there are three algorithm parameters to be defined by user, which is maximum iteration times t , population size s and the number of complexes p .

3 Study area and data used

The ice breakup prediction in the Bayangaole gauging station was performed in this study, the gauging station located in the Inner Mongolia section of the Yellow River, as shown in Fig. 2. The Inner Mongolia reach lies in the top north of the Yellow River, with the altitude of more than 1000 m, it has a cold and long winter. Usually, the Inner Mongolia reach has about the 4–5 months ice period every year. Due to the variation of the time and strength of the cold wave intrusion and the influence of wind, annual ice flood date always varied greatly. The ice-break often occurs in the second and last ten days of March, and seldom occurs in the first ten days of April. During every spring, ice floods occur frequently and bring huge loss, especially with the development of the social economy, the loss will be greater.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



The breakup date are calculated from the reference date (1 May), for example, breakup date of 46 days means that the river ice break up 46 days before 1 May, i.e. on 15 March. By using the correlation analysis, three factors are selected as the forecast factors, that is, accumulated positive air temperature from the date that temperature goes above zero to the break-up date, average water level and average streamflow of break-up period. Data used for ice breakup forecast was collected by Ji (2002). The first 29 samples are used for training and the rest 5 samples for validation.

4 Application of SVM to ice breakup predictor

4.1 Prediction model

The implementation of SVM for forecasting ice breakup can be generalized as the following form:

$$Y = f(X|\theta), \tag{9}$$

where Y is $N \times 1$ vector of model predictions, X is $N \times 3$ matrix of input variables and θ is a vector of n unknown SVM parameters. In this case, Y is the annual ice breakup date, X including the following input variants: accumulated daily positive air temperature, average water level and average streamflow, θ refers to two SVM parameters (C and ε) and kernel function parameters.

The most widely used kernel function is the Gaussian radical basis function (RBF) as follows, which has one parameter σ :

$$K(x_j, x) = \exp\left(\frac{-0.5\|x - x_j\|^2}{\sigma^2}\right). \tag{10}$$

Once the suitable SVM parameters are selected, training should be conducted for forecasting. The SVM software developed by Steve Gunn (2001) was used in this study.

4.2 Data preprocessing

Before training the SVM, data (both the input and output data) should be preprocessed for the two reasons. Firstly, preprocessing can ensure that all variables receive equal attention during the training process. Secondly, preprocessing is important for the efficiency of training algorithms (Wang et al., 2006). In general, there are two types of preprocessing methods. The first one is to rescale the data to a small interval, for example, $[-1, 1]$, $[-0.9, 0.9]$, $[0.1, 0.9]$ or $[0, 1]$. Another is to standardize the data by subtracting the mean value and then dividing by the standard deviation, that is rescale the data to a Gaussian function with a mean of 0 and standard deviation of 1. The advantage of using rescaled interval of $[0.1, 0.9]$ is that extreme events occurring outside the range of the calibration data may be accommodated (Dawson and Wilby, 1999). In this study, all data is rescaled to the interval of $[0.1, 0.9]$ using the equation

$$x'_i = 0.8 \cdot \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} + 0.1, \quad (11)$$

where x'_i is the scaled value, x_i is the original value, x_{\min} , x_{\max} is the minimal and maximum value of x , respectively.

4.3 Performance measures

There are many literatures on model forecasting evaluating indices. The following numerical performance statistics that are defined here are used to evaluate the forecasting results.

1. Mean absolute error (MAE)

$$\text{MAE} = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n}. \quad (12)$$

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



2. Root mean squared error (RMSE)

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}. \quad (13)$$

3. Correlated coefficient (CC)

$$\text{CC} = \frac{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y}) (\hat{y}_i - \bar{\hat{y}})}{\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2} \cdot \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})^2}}. \quad (14)$$

4. Mean relative error (MRE)

$$\text{MRE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|. \quad (15)$$

In Eqs. (12)–(15), y_i is the observed value, \bar{y} is the mean observed value, \hat{y}_i is the predicted value, $\bar{\hat{y}}$ is the mean predicted value, n is the total number of predicted values.

4.4 Prediction steps

Procedure for ice breakup forecasting can be summarized as the following:

- (1) Setting parameters of MOSCEM algorithm. In this study, t , s and p use the value of 1000, 5000 and 10, respectively.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



(2) Define the SVM parameter search space. Gaussian kernel function was employed in this paper, so there are three parameters to be optimized, C , ε , σ . Hsu et al. (2003) suggested trying exponentially growing sequences of C and σ is a practical method to identify good parameters, for example, $C \in [2^{-5}, 2^5]$. Search space of the parameters in this study is $C \in [2^0, 2^{10}]$, $\varepsilon \in [0, 0.5]$, $\sigma \in [2^{-10}, 2^0]$.

(3) SVM training using MOSCEM algorithm. Two objective functions were considered:

$$\min F1(\theta) = \text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}, \quad (16)$$

$$\max F2(\theta) = \text{CC} = \frac{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2} \cdot \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})^2}}, \quad (17)$$

where n is the number of testing samples.

The MOSCEM algorithm software developed by Vrugt et al. (2003) was used in this study. For this case, the three parameters value was selected through running the MOSCEM algorithm, the optimal values are $C=75.35$, $\varepsilon=0.0375$, $\sigma=0.2199$.

(4) Prediction using the SVM with optimal parameters.

4.5 Comparing model

Artificial neural network (ANN) was also used for comparing in this study. The key problem for constructing the ANN is to determine the number of hidden layer neurons and training algorithm. Three layers of ANN is constructed, input and output variables are same to the SVM. Tan-sigmoid transfer function and linear transfer function were

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



employed for hidden layer and output layer, respectively. Training epoch is set to 500. To prevent overtraining and improve the generalization ability, Bayesian regularization method was used for training the ANN. No unified theory exists for determining the number of hidden layer neurons. A trial and error procedure is generally used to determine the optimal numbers. In this study, four hidden layer neurons was selected that yielding the best performance.

4.6 Results and discussions

To facilitate comparison, same training and testing samples are used for SVM and ANN. Figure 3 depicts the training and testing performance of the two models. The scatter plots of observed versus forecast day of the ice breakup are shown in Fig. 4. Forecast results are listed in Tables 2 and 3 compares the performance indices obtained with the forecasting results of the SVM and ANN. It is show that SVM outperforms ANN considerably both in training period and testing period, it also proved that SVM is more suitable for small samples prediction than ANN. Figure 4 shows that the errors of training and testing period of SVM are smaller than that of ANN.

The major drawback of ANN is the fact that it do not lead to one global or unique solution compared with the SVM. It is mainly due to the differences of the initial weights of the ANN. Furthermore, the SVM training process always seeks a global optimized solution and avoids overfitting that eventually leads to better generalization performance than neural network models. Despite its superior features, SVM also have some limitations.

- (1) In order to construct an efficient SVM model, hyper-parameters must be selected properly. Otherwise, it may lead to over-fitting or under-fitting. Different parameter sets may make a great difference in performance. In this study, MOSCEM algorithm is used for hyper-parameters selecting, the forecast result shows the effectiveness of the algorithm. But experiences shows that in order to find the optimal hyper-parameter set, the number of random samples should be set to

The support vector machines approach

H. Zhou et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



The support vector machines approach

H. Zhou et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



a large number, when the number of training samples is very large, it will be very time consuming. Therefore, the effectiveness and efficiency of the MOSCEM algorithm depends on the parameters of its own. So in order to get an ideal performance, these parameters of MOSCEM should be determined by trial and error.

(2) Many functions can be used as the kernel function for SVM, as Table 1 showed. In this study, only radial basis kernel function is used, and no comparison is made for the performances of different kernel function for SVM. It should be the future research work.

(3) In this study, due to the limitation of data, there are only 34 samples for training and validation. As the number increasing, the model should be further investigated.

5 Conclusions

In this study, SVM is employed for ice breakup prediction in Bayangaole gauging station. To build reliable forecasting models, MOSCEM algorithm is implemented for selecting optimal parameters for SVM in a multiobjective framework. This multi-criteria method can be used to determine the optimal model structure. Compared with the ANN model, SVM model appears to be more suitable for ice breakup forecast. Due to the limitation of the data set, only 34 samples are used for forecast in this work, as the increasing of data set number, the proposed model based on SVM should be further investigated in the ongoing study.

References

ASCE: Artificial neural networks in hydrology. I: Preliminary concepts, J. Hydrol. Eng., 5(2), 115–123, 2000.

- ASCE: Artificial neural networks in hydrology. II: Hydrologic applications, *J. Hydrol. Eng.*, 5(2), 124–137, 2000.
- Beltaos, S.: Advances in river ice hydrology, *Hydrol. Process.*, 14(9), 1613–1625, 2000.
- Beltaos, S.: Numerical modelling of ice-jam flooding on the Peace-Athabasca delta, *Hydrol. Process.*, 17(18), 3685–3702, 2003.
- Chen, S. Y. and Ji, H. L.: Fuzzy optimization neural network approach for ice forecast in the Inner Mongolia reach of the Yellow River, *Hydrolog. Sci. J.*, 50(2), 319–330, 2005.
- Dawson, C. W. and Wilby, R. L.: A comparison of artificial neural networks used for river forecasting, *Hydrol. Earth Syst. Sci.*, 3(4), 529–540, 1999.
- Duan, Q. , Gupta, V. K., and Sorooshian, S.: Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resour. Res.*, 28, 1015–1031, 1992.
- Hsu, C. W., Chang, C. C., Lin, C. J.: A practical guide to support vector classification. Technical report, Department of Computer Science and Information Engineering, National Taiwan University, Taiwan, 2003.
- Huang, W., Nakamori, Y., Wang, S. Y., et al.: Forecasting stock market movement direction with support vector machine, *Comput. Oper. Res.*, 32(10), 2513–2522, 2005.
- Ji, H. L.: Factor analysis for ice flood and model research for freeze-up date and break-up date in the Inner Mongolia Reach of the Yellow River. Doctorate Dissertation, College of Water Conservancy and Civil Engineering, Inner Mongolia Agricultural University, 2002 (in Chinese).
- Mahabir, C., Hicks, F., Fayek, A. R., et al.: Neuro-fuzzy river ice breakup forecasting system, *Cold Reg. Sci. Technol.*, 46(2), 100–112, 2006.
- Massie, D. D., White, K. D., Daly, S. F., et al.: Application of neural networks to predict ice jam occurrence, *Cold Reg. Sci. Technol.*, 35(2), 115–122, 2002.
- Morse, B. and Hicks, F.: Advances in river ice hydrology 1999–2003, *Hydrol. Process.*, 19(1), 247–263, 2005.
- Niu, D. X., Wang, Q., Li, J. C., et al.: Short Term Load Forecasting Model Based on Support Vector Machine, in: *Advances in Machine Learning and Cybernetics*, vol. 3930, Springer-Verlag, Berlin, 880–888, 2006.
- Gunn, S.: <http://www.isis.ecs.soton.ac.uk/resources/svminfo/>, 2001.
- Sun, W. and Yang, C. G.: Research of Least Square Support Vector Machine Based on Chaotic Time Series in Power Load Forecasting Model. in: *Neural Information Processing*, part 2, Proceedings, vol. 4233, Springer-Verlag, Berlin, 984–993, 2006.

The support vector machines approach

H. Zhou et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



- Vapnik, V.: The Nature of Statistical Learning Theory, Springer, New York, 1995.
- Vapnik, V.: Statistical Learning Theory, John Wiley and Sons, New York, 1998.
- Vrugt, J. A., Gupta, H. V., Bastidas, L. A., et al.: Effective and efficient algorithm for multiobjective optimization of hydrologic models, *Water Resour. Res.*, 39(8), p. 19, 2003.
- 5 Wang, W., Van Gelder, P., Vrijling, J. K., et al.: Forecasting daily streamflow using hybrid ANN models, *J. Hydrol.*, 324(1–4): 383–399, 2006.
- White, K. D.: Predicting breakup ice jams using logistic regression, *J. Cold Reg. Eng.*, 10(4), 178–189, 1996.
- White, K. D.: Review of prediction methods for breakup ice jams, *Can. J. Civil Eng.*, 30(1), 89–100, 2003.
- 10 Ye, M. Y., Wang, X. D., Zhang, H. R., et al.: Chaotic time series forecasting using online least squares support vector machine regression, *Acta Phys. Sin.-Ch. Ed.*, 54(6), 2568–2573, 2005.

HESSD

6, 3175–3198, 2009

The support vector machines approach

H. Zhou et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



The support vector machines approach

H. Zhou et al.

Table 1. Kernel functions for SVR.

Kernels	Functions	Parameters
Linear	$\langle x, x_i \rangle$	
Polynomial	$(\langle x, x_i \rangle + 1)^d$	d
Radial basis function	$\exp\left(-\ x - x_i\ ^2 / 2\sigma^2\right)$	σ
Sigmoid	$\tanh(b\langle x, x_i \rangle + c)$	b, c

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



The support vector machines approach

H. Zhou et al.

Table 2. Forecasting results of the SVM and ANN.

No.	Year	Observed value (d)	Forecast value (d)		Error (d)		MAE (d)	
			ANN	SVM	ANN	SVM	ANN	SVM
1	1997–1998	41	44	42	3	1	2.6	1.8
2	1998–1999	51	53	51	2	0		
3	1999–2000	54	53	51	1	3		
4	2000–2001	54	59	51	5	3		
5	2001–2002	45	47	43	2	2		

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



The support vector machines approach

H. Zhou et al.

Table 3. Comparison of the forecasting results from each model.

Model	Training period			Testing period		
	RMSE	CC	MRE (%)	RMSE	CC	MRE (%)
ANN	6.6	0.84	12.8	2.95	0.93	10.9
SVM	3.82	0.95	6.8	2.14	0.96	3.6

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



The support vector machines approach

H. Zhou et al.

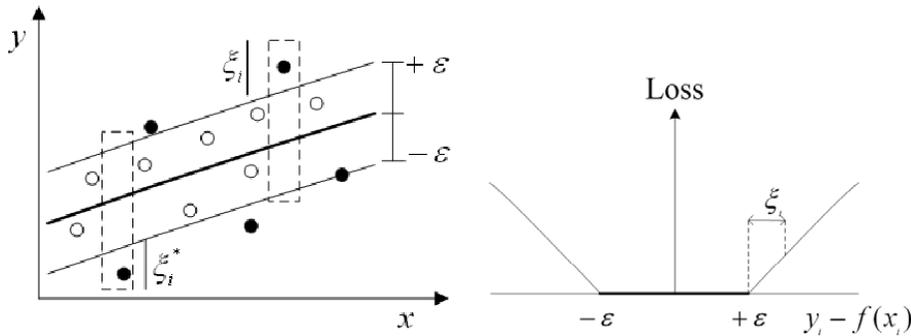


Fig. 1. Soft margin loss setting for a linear SVR.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



The support vector machines approach

H. Zhou et al.

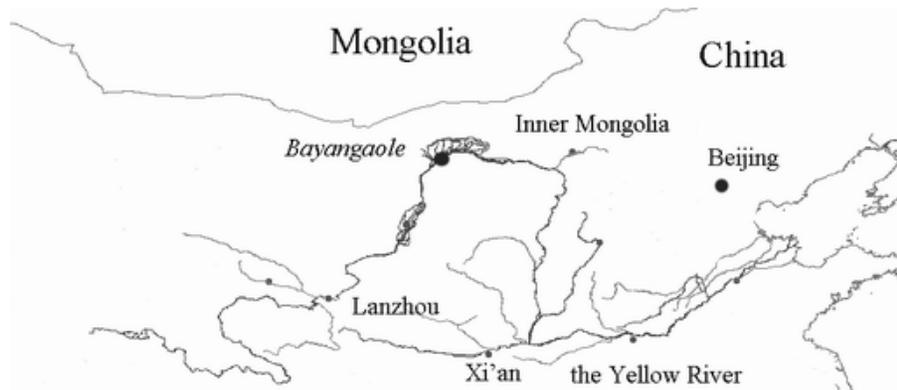


Fig. 2. The location of Bayangaole gauging station.

[Title Page](#)

[Abstract](#)

[Introduction](#)

[Conclusions](#)

[References](#)

[Tables](#)

[Figures](#)

[I◀](#)

[▶I](#)

[◀](#)

[▶](#)

[Back](#)

[Close](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)



The support vector machines approach

H. Zhou et al.

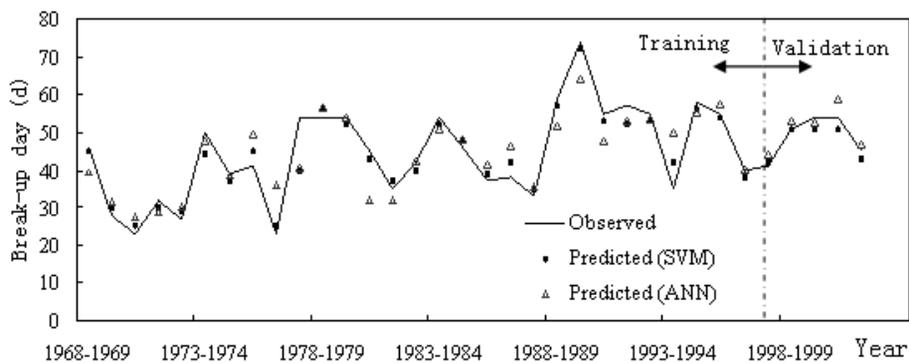


Fig. 3. Models performances for breakup prediction during training and testing period.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



The support vector machines approach

H. Zhou et al.

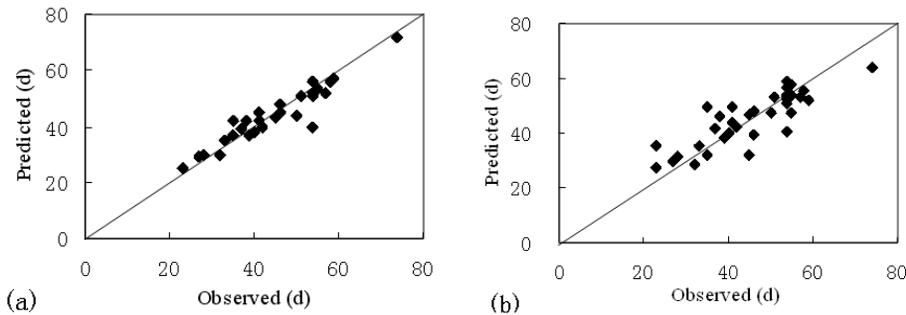


Fig. 4. Scatter plots of forecast versus observed ice breakup date for the training and testing data sets: **(a)** SVM **(b)** ANN.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

