Developing a Proper Model for Online Estimation of the 5-Day Biochemical Oxygen Demand Based on Artificial Neural Network and Support Vector Machine

Eskandari, A.[*1], Nouri, R.[2], Meraji, S.H.[3], Kiaghadi, A.[4]

1- M.Sc., Hydraulic Engineering, Department of Civil Engineering, Faculty of Engineering, Islamic Azad University, Bushehr - Iran

2- Ph.D. student, Environmental Engineering, Invited Lecturer of Civil Engineering, Faculty of Engineering, Islamic Azad University, Bushehr - Iran roohollahnoori@gmail.com

3- Ph.D. Student, Hydraulic Engineering, Invited Lecturer of Civil Engineering, Faculty of Engineering, Islamic Azad University, Bushehr - Iran hamedmeraji@gmail.com

4- M.Sc. Student, Environmental Engineering, Invited Lecturer of Civil Engineering, Faculty of Engineering, Islamic Azad University, Bushehr - Iran amin.kiaghadi@gmail.com

Received: March, 2011 Accepted: June, 2011

Introduction

Five-day biochemical oxygen demand (BOD_5) is one of the key parameters which is widely applied to evaluate the biological and chemical conditions of rivers, and to study the effects of releases in rivers from wastewater treatment plants, factories, farms, etc. Measurement of this parameter requires about 5 days. However, 5 days is a long time for the determination of BOD_5. Therefore, in the present research two software sensors, i.e. artificial neural network (ANN) and support vector machine (SVM) are developed for the online estimation of BOD_5. To achieve this goal, the Sefidrood River in Iran was selected. It is pointed out that although ANN models have been used to predict BOD_5, but using SVM techniques for the online estimation of this parameter have never been reported.

Materials and Methods

ANN and SVM models are two well known artificial intelligent techniques which have frequently been used for modeling nonlinear parameters in engineering problems. In this research, a feed-forward ANN technique and a ε-SVM regression model is considered for online prediction of BOD_5. To optimize ANN parameters, three optimization algorithms, i.e. Levenberg-Marquardt (LM), resilient back-propagation (RP), and scaled conjugate gradient (SCG) are used. Besides, two-steps grid search algorithm is considered for the optimization of SVM parameters, i.e. ε, C and γ.

Results and Discussion

To achieve the best network geometry according to the highest correlation (R), and the lowest root mean square error (RMSE) of the testing sets for online estimation of BOD_5, various structures of NN model with different training functions are investigated. Finally, the network with LM algorithm (ANN (LM) model) is selected as the best ANN model. The results of training and testing steps for developed ANN models with different training functions are given in Table 1.

In addition, for SVM model development, first, ε, C and γ should be determined. In this research, the best fitting γ value was obtained by trial and error procedure. Therefore, this parameter was set to several values. Thus, 14 models were developed (ε and *C* are considered to be varied). The RMSE value has been assessed on different γ values in the different ε-SVM regression models. Obtained results indicated that the SVM model including the γ value equal to 1.472 has the least RMSE.

Therefore, it can be chosen as the best SVM model for the online prediction of BOD_5. However, optimal values for *C* and ε using two-step grid search were 13 and 0.037, respectively. Finally, SVM model is trained and tested using the determined best fitted values.

Corresponding author: Tel: 09177710039 E-mail: ali_eskandari225@yahoo.com

Table 1 illustrates training and testing results of developed SVM model.

Table 1: Obtained results for ANN and SVM models in the training and testing steps

Model	Training function	Training		Testing	
		R	RMSE	R	RMES
ANN	LM	0.93	7.5	0.90	11.6
	RP	0.90	7.9	0.87	12.1
	SCG	0.88	8.2	0.85	12.8
SVM	Two-steps grid search	0.97	4.6	0.95	9.16

Finally, for evaluating the performance of the selected developed models (SVM and ANN (LM) models), the developed discrepancy ratio (DDR) statistic is used. Fig. 1 shows results of DDR statistic in the testing step of SVM and ANN (LM) models.

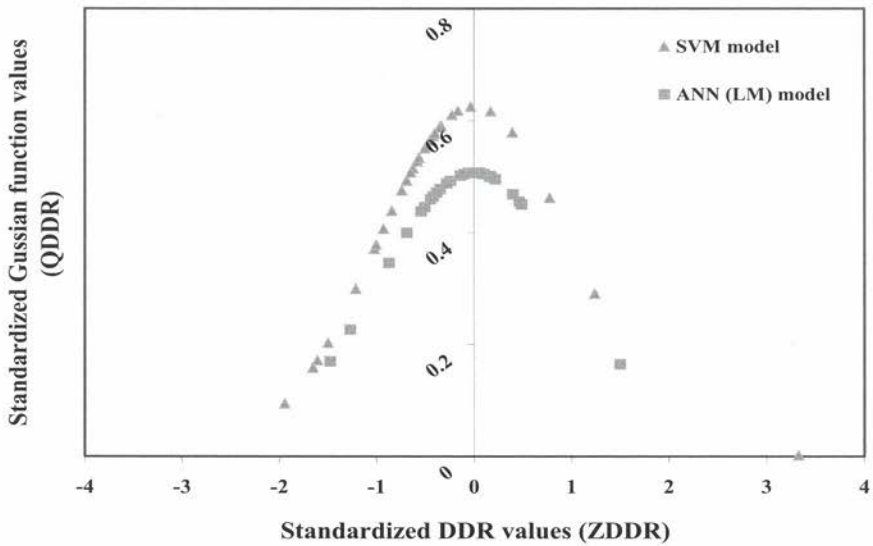

Fig. 1: Standardized normal distribution graph of DDR values for testing steps of ANN (LM) and SVM models

It should be noted that in the obtained figure, more tendencies to the centerline and also, a larger value of the maximum normalized value for DDR (Q_{DDR}) indicate more accuracy. This figure indicates that SVM model has more tendencies to the centerline than ANN (LM) model. In addition, maximum Q_{DDR} of the SVM model (0.62) is larger than that of the ANN (LM) model (0.51).

Therefore, results of DDR analysis reveal that SVM model is superior. According to the obtained results from DDR analysis, the SVM model is selected for the online prediction of BOD_5 in Sefidrood River, Iran. Training and testing results of the SVM model is shown in Fig. 2.

Conclusion

Recently, hardware sensors are widely used in monitoring and measurement of water quality parameters. Constraint of the instrument to measure some water quality parameters such as BOD_5, which are time consuming, causes efforts to be directed to the use of software sensors for the online prediction of BOD_5. The main goal of this research was to develop an appropriate software sensor based on artificial neural network (ANN) and supported vector machines (SVM) models for the online prediction of BOD_5 in the Sefidrood River.

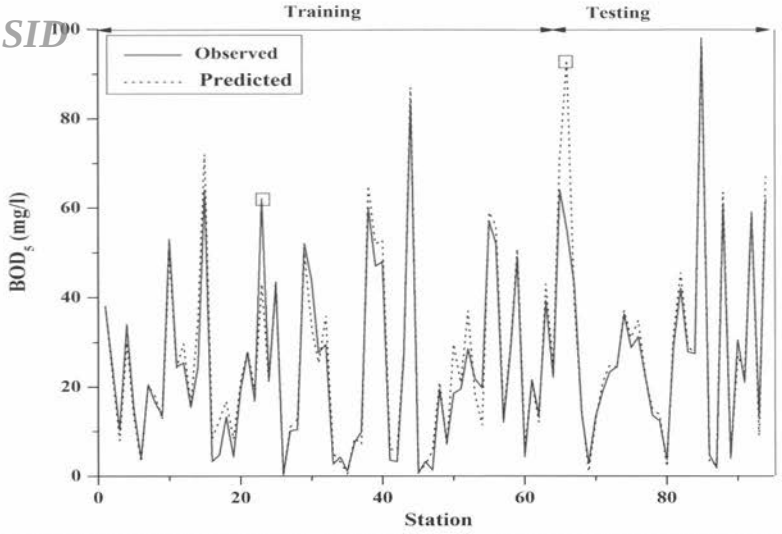

Fig. 2: Predicted and observed BOD_5 for training and testing stages of SVM model
□Indicates the extreme high values which is resulted from the model with some errors.

For this purpose, appropriate models with ANN and SVM are developed by considering BOD_5 as a function of other water quality variables. In the development of ANN model, the role of various training functions such as LM, resilient back-propagation, RP and SCG algorithms on the optimization of ANN parameters is evaluated. Also for the optimization of SVM parameters, two-step grid search algorithm is conducted. The results of this research indicated the superior performance of ANN (LM) model In comparison with the other two algorithms, i.e. RP and SCG. Besides, SVM model had a suitable performance in BOD_5 prediction. Finally, further investigations to select the best model between ANN (LM) and SVM based on the developed discrepancy ratio statistics revealed that SVM model was superior.

Key words
Artificial Neural Network (ANN), Support Vector Machine (SVM), Sefidrood River, 5-Day Biochemical Oxygen Demand (BOD_5)