Using ensemble model to improve ANN, ANFIS, SVR models in predicting effluent BOD and COD

Parisa Asghari[1], Vahid Nourani[2*], Elnaz Sharghi[3], Nazanin Behfar[4]

[1] M.Sc., Department of Water Resources Engineering, Faculty of Civil Engineering, University of Tabriz, Tabriz, Iran

[2] Professor, Department of Water Resources Engineering, Faculty of Civil Engineering, University of Tabriz, Tabriz, Iran

[3] Assistant Professor, Department of Water Resources Engineering, Faculty of Civil Engineering, University of Tabriz, Tabriz, Iran

[4] PhD student, Department of Water Resources Engineering, Faculty of Civil Engineering, University of Tabriz, Tabriz, Iran

ABSTRACT

In this study, black box artificial intelligence models (AI) including feed forward neural network (FFNN), support vector regression (SVR) and adaptive neuro-fuzzy inference system (ANFIS) were used to predict effluent biological oxygen demand (BOD_{eff}) and chemical oxygen demand (COD_{eff}) of Tabriz wastewater treatment plant (WWTP) using the daily data collected from 2016 to 2018. In addition, the autoregressive integrated moving average (ARIMA) linear model was used to predict BOD_{eff} and COD_{eff} parameters in order to compare the linear and non-linear models abilities in complex processes prediction. To improve the prediction of BOD_{eff} and COD_{eff} parameters, the data post-processing ensemble method was also used. The input data set included daily influent BOD, COD, total suspended solids (TSS), pH at the current time (t) and BOD_{eff} and COD_{eff} at the previous time (t-1) and the output data included BOD_{eff} and COD_{eff} at t. The results of the single models indicated that SVR model provides better results than the other single models. In the ensemble modeling, simple and weighted linear averaging, and neural network ensemble methods were applied to enhance the performance of the single AI models. The results indicated that using ensemble models could increase the prediction accuracy up to 15% at the verification phase.

KEYWORDS

Soft Computing, Artificial Intelligence, ARIMA Linear Model, Ensemble, Wastewater Treatment Plant

* Corresponding Author: Email: nourani@tabrizu.ac.ir

1. Introduction

Wastewater treatment plants (WWTPs) have complex physical, biological, and chemical processes usually with nonlinear behavior. Therefore, predicting such complex processes by physical-based models is somewhat costly and time consuming. So, to increase the accuracy and improve the operation of WWTPs, the use of artificial intelligence (AI) methods is highly recommended [1]. In this regards, researchers have used AI methods to investigate complex phenomena such as wastewater treatment processes. For examples, Mjalli et al. [2] used the ANN model to predict Doha West WWTP performance. Sherif et al. [3] evaluated the performance of SVR and ANFIS models to predict the performance of aerobic granule sludge reactors and showed that the SVR model provides more accurate predictions than the ANFIS model.

Each of the AI models due to its unique structure or features may have failures or errors in modeling a phenomenon. Recently, a model combination approach to reduce model errors and achieve a comprehensive modeling has been presented as models ensemble framework (see [4-8]). In this study, the outputs of several models were combined via three ensemble techniques i.e. simple averaging ensemble (SAE), weighted averaging ensemble (WAE) and neural network ensemble (NNE) to enhance the overall prediction of BOD_{eff} and COD_{eff}.

Considering the aforementioned studies, as a novelty of this paper ensemble model was employed to improve efficiency of BOD_{eff} and COD_{eff} modeling. Moreover, BOD^{t-1}_{eff} and COD^{t-1}_{eff} were used as input variables for the first time.

2. Materials and methods

2.1. Case study and Data

Tabriz is one of the urbanized and industrialized cities of Iran. In this study, daily data were gathered from Tabriz WWTP over 2 years (2016-2018). Considering the significance role of BOD and COD parameters in effluent quality control, these parameters were selected separately as the models outputs, thus, each model has a single output. The correlation coefficient (CC) was utilized to determine the effect of each influent parameter on the BOD and COD, so, appropriate inputs were selected with a CC value greater than or equal to 0.20 for the input layer of the models.

2.2. Feed Forward Neural Network (FFNN)

ANN is used as a method to predict nonlinear phenomena. The FFNN method and the back propagation (BP) algorithm are widely used in various fields of science and they have presented acceptable predictions at various fields of environmental engineering. The ANN structure consists of three layers: input, hidden and output layer.

2.3. Adaptive Neural Fuzzy Inference System (ANFIS)

ANFIS is a hybrid model of fuzzy logic and neural network concepts to use the advantages of both methods as a unique model. Each fuzzy system consists of three main phases: fuzzification, fuzzy databases and defuzzification.

2.4. Support Vector Regression (SVR)

Support vector regression developed on the basis of the support vector machine (SVM) fundamentals. These neural networks are one of the supervised learning methods and unlike other neural networks, they consider operational risk as the target to be optimized instead of minimizing the computational errors between obtained and observed values.

2.5. Auto-Regressive Integrated Moving Average (ARIMA)

The ARIMA model is an extended ARMA (Auto Regressive Moving Average) model that is used for time series that are not static. ARIMA is a linear model and it is capable of identifying the patterns of data and it allows for predicting the future based on past input data.

2.6. Ensemble of outputs

The main idea behind this multi-model approach is to use the unique capabilities of each model to estimate the data series pattern better. In this study, the ensemble model was performed employing three simple and weighted linear averaging and neural network ensemble techniques.

3. Results and Discussion

At the first of modeling, FFNN, ANFIS, SVR, and ARIMA models were applied to predict BOD_{eff} and COD_{eff}. The results of modeling are briefly presented in Table 1.

As shown in Table1, the SVR model has presented better results than the other models. In the following, the results of four single models are merged by three methods of ensemble models (i.e. SAE, WAE, and NNE). The results of ensemble model is tabulated in Table 2.

Table 1. Results of BOD_{eff} and COD_{eff} predictions by single models of FFNN, ANFIS, SVR and ARIMA.

Model	Parameter	DC	
		Train	Validation
FFNN	BOD	0.8388	0.7182
	COD	0.8633	0.7178
ANFIS	BOD	0.8451	0.7203
	COD	0.8440	0.7148
SVR	BOD	0.8521	0.7423
	COD	0.8673	0.7119
ARIMA	BOD	0.8435	0.5281
	COD	0.8339	0.6279

Table 2. Ensemble modeling results to predict BOD_{eff} and COD_{eff}.

Parameter	Ensemble Model	DC	
		Train	Validation
BOD_{eff}	SAE	0.8473	0.7972
	WAE	0.8498	0.7972
	NNE	0.8687	0.8120
COD_{eff}	SAE	0.8786	0.8193
	WAE	0.8818	0.8198
	NNE	0.9238	0.8281

According to Table 2, using ensemble model increased the modeling accuracy. This is because integrating outputs of individual models, decreases variance, and bias and improves the modeling efficiency. The SAE, WAE and NNE methods increased AI modeling performance for BOD_{eff} and COD_{eff} parameters by 11%, 11%, 13% and 14%, 14%, 15% in validation phases over the single FFNN model, respectively.

4. Conclusion

In this study, BOD_{eff} and COD_{eff} parameters of Tabriz WWTP were predicted via ANFIS, FFNN, SVR, and ARIMA models. Then, these models were combined with three techniques of ensemble model. The following key facts are obtained:

1) ARIMA model due to its linearity cannot predict complex phenomena such as wastewater treatment plants.
2) SVR model was provided more accurate results than other single models.
3) Ensemble models provided a better approximation than single models. Among three different ensemble techniques, the NNE model was more efficient and it could increase the accuracy of daily BOD_{eff} and COD_{eff} modeling performance up to 13% and 15% in validation steps, respectively. Due to the nonlinear nature of the WWTP processes, the use of nonlinear ensemble method (NNE) is recommended.

5. References

[1] R.S. Govindaraju, Artificial neural networks in hydrology. II: hydrologic applications, Journal of Hydrologic Engineering, 5(2) (2000) 124-137.

[2] F.S. Mjalli, S. Al-Asheh, H. Alfadala, Use of artificial neural network black-box modeling for the prediction of wastewater treatment plants performance, Journal of Environmental Management, 83(3) (2007) 329-338.

[3] M.S. Zaghloul, R.A. Hamza, O.T. Iorhemen, J.H. Tay, Comparison of adaptive neuro-fuzzy inference systems (ANFIS) and support vector regression (SVR) for data-driven modelling of aerobic granular sludge reactors, Journal of Environmental Chemical Engineering, 8(3) (2020) 103742.

[4] J.M. Bates, C.W. Granger, The combination of forecasts, Journal of the Operational Research Society, 20(4) (1969) 451-468.

[5] R.T. Clemen, Combining forecasts: A review and annotated bibliography, International journal of forecasting, 5(4) (1989) 559-583.

[6] J. Dickinson, Some comments on the combination of forecasts, Journal of the Operational Research Society, 26(1) (1975) 205-210.

[7] E. Sharghi, V. Nourani, N. Behfar, Earthfill dam seepage analysis using ensemble artificial intelligence based modeling, Journal of Hydroinformatics, 20(5) (2018) 1071-1084.

[8] P.D. Thompson, How to improve accuracy by combining independent forecasts, Monthly Weather Review, 105(2) (1977) 228-229.