Ensemble of ANN and ANFIS for Water Quality Prediction and Analysis - A Data Driven Approach

Y.Khan, S.S.Chai

Faculty of Computer Science and Information Technology,
Universiti Malaysia Sarawak, Kota Samarahan 94300, Malaysia.
yafra.khan@gmail.com

Abstract—The consequences of un-clean water are some of the direst issues faced by humanity today. These concerns can be addressed efficiently if data is pre-analyzed and water quality is predicted before its effects occur. The aim of this research is to develop a novel ensemble of Artificial Neural Network (ANN) and Adaptive Neuro-Fuzzy Inference System (ANFIS) models using averaging ensemble technique, producing improved prediction accuracy. Measurements of different water quality parameters have been used for predicting the overall water quality, applying ANN, ANFIS and ANN-ANFIS ensemble and their results have been compared. The data used in this study is obtained by USGS online repository for the year of 2015, with a 30-minutes time interval between measurements. Root Mean Squared Error (RMSE) has been used as the main performance measure. The results depict a significant improvement in the Ensemble ANN-ANFIS model (RMSE: 0.457) as compared to both the ANN model (RMSE: 2.709) and the ANFIS model (1.734). The study concludes that the ensemble of ANN and ANFIS model shows significant improvement in prediction performance as compared to the individual models. The research can prove to be beneficial for decision making in terms of water quality improvement.

Index Terms—Water Quality Prediction; Artificial Neural Networks; Adaptive Neuro-Fuzzy Inference System; Ensemble Learning; Machine Learning.

I. INTRODUCTION

The contamination of natural water resources is quite rampant due to its wide availability. This contamination is the result of various factors including poor sanitation infrastructure and lack of awareness [1]. This engenders a dire need for adopting innovative approaches and techniques for water quality prediction before its consequences arise and take the precautionary actions. Water quality can be evaluated by either a single parameter for a specific use or by multiple Water Quality (WQ) parameters. In case of multiple WQ parameters, a Water Quality Index (WQI) is used, which is a numerical representation of the quality of a water resource covering various significant water quality parameters in connection with a set of water quality standards [2].

For carrying out the analysis and prediction of water quality, various studies have proposed and implemented different methodologies [3]. One such study [4] proposes Reasoning Based Expert System (RBES) to compare the water quality parameters with the industry standards from the knowledge base to make a decision. Besides that, time-series analysis techniques like Auto-Regressive Integrated Moving Average (ARIMA) have been widely used in this regard [4][6]. More recently, Support Vector Machine (SVM) has been applied in water quality prediction scenario [7] to predict the concentration of one parameter in water based upon the values of other water quality parameters.

Despite improving results in the above mentioned techniques, following few points need to be considered when selecting a suitable technique for water quality prediction: a) Mapping input-output data in case of water quality dataset becomes very complex due to non-linear nature of water quality dataset with linear modeling approaches [8] [9] (b) Prediction accuracy and model simplicity needs to be considered (c) Simplified interpretation of input-output relationship in order to deal with uncertainties (d) Combining multiple models improves generalization and diversity of the model.

Artificial Neural Network (ANN) is one technique that has proved to be effective in not only describing nonlinear input-output relationship of complex datasets, but also in providing strong model flexibility [10]. ANN has been applied successfully for other complex prediction scenarios like groundwater level prediction [11] and wind speed forecasting [12]. In case of water quality prediction, Gazzaz et.al. [2] use Multi-Layer Perceptron (MLP) Neural Network to predict WQI based upon certain WQ parameters. The result in terms of RMSE turns out to be effective. On the downside, ANN is a black-box approach, hence the model simplicity is compromised and uncertainty is not effectively dealt with. Adaptive Neuro-Fuzzy Inference System (ANFIS) has been found to be an effective approach in this regard, using the interpretability aspect of fuzzy inference while retaining the benefits of ANN [11]. ANFIS has been found to be suitable in modeling complex datasets like that of hydrological applications. One such study applied ANFIS for prediction of oily wastewater microfiltration permeate volume and was found to be a reliable approach [13]. Similarly, Talebizadeh and Moridnejad [14] carried out a comparison of ANN and ANFIS in forecasting lake level fluctuations, where ANFIS turned out to be superior than ANN in terms of efficiency. In case of water quality prediction, ANFIS has been applied for Biochemical Oxygen Demand (BOD) prediction based upon other WQ parameters as inputs [15]. This study shows a significant accuracy for different input combinations, with MSE between 1.2 and 2.5.

Despite prediction model improvements, increase in model accuracy while avoiding over-fitting is still a challenge for most researchers. According to recent researches, model performance can be significantly improved if an appropriate hybrid of multiple models is used for forecasting and prediction than using a single model in this regard [16][6]. Ensemble learning refers to a process of combining multiple predictors in order to boost the model performance. It uses a combination or a committee of relatively "weak" learners to