MACHINE LEARNING TECHNIQUES FOR WASTEWATER TREATMENT PLANTS

***Kirthana T[1], Soumya A[2]**

[1,2]*Department of Computer Science & Engineering , RV College of Engineering, Bengaluru, Karnataka, India*

[1] *kirthanat.scs18@rvce.edu.in,* [2]*soumysaa@rvce.edu.in*

Abstract: *Wastewater Treatment is a relevant topic in present days. Treating wastewater and make it reusable is a challenging task. Because of growing amount of integrated food waste, wastewater treatment plants are constructed and that necessitated an efficient method to perform treatment. Currently, there are models which are generated using linear, logarithmic, and exponential functions. Since the variables are related in a non-linear manner, these functions might not suffice the performance expectation. Machine learning techniques are applied to obtain better results. In this paper, a comprehensive review is done on different Machine Learning approaches and algorithms, implemented for wastewater treatment.*

Keywords: **Wastewater Treatment Plants (WWTPs), Linear function, Logarithmic function, Exponential function, Machine learning.**

1. Introduction

The major priority for a green environment is to enhance the quality of water and efficiency in wastewater treatment. Since water is a major source of living, the reuse of treated water is a need in many countries. Water resources should be protected along with financial benefits. Hence, recycling of wastewater, a cost-effective process is encouraged than desalting seawater. However, the energy required in managing wastewater treatment plants is high. This necessitates adoption of potential methods to save cost and energy. Wastewater treatment involves traditional methods for separating solid waste which can be easily performed, but dissolved components need to be separated chemically. Processing of water accounts nearly 2-3% of whole world's electricity consumption. Different factors that impact the wastewater treatment is analyzed and different strategies are studied.

Several cost models are developed to evaluate the relationship among the process variables and the cost of Wastewater Treatment Plant using traditional regression techniques. A study on various dominant approaches to data- based modeling is carried out in the paper.

2. Fundamentals of Machine Learning Techniques

The fundamental environmental treatment facilities like Wastewater Treatment Plants (WWTPs) are need of days. In the inland water treatment, the solid wastes are physically separated, and remaining wastes are separated by recycling in wastewater treatment plant.

Several cost models are proposed to assess the relationship between the most relevant process variables and the cost of WWTP using traditional regression techniques like linear, logistic, and exponential regression.

Machine learning is a process of training the machine with the input and output. The algorithm is defined to learn from the trained examples and predict the result for new inputs. Supervised Learning technique predicts the outcome of unknown data by applying some function on the labeled training data. The input parameters for the functions play a fundamental role in choosing the technique. The output variables are grouped to classes if the values are discrete. Under unsupervised learning, clustering is done when the values of the parameters are close enough to get

clustered. In Reinforcement learning (RL), a keen observation is done by the agent on environment by continuously interacting with it.

Neural Networks (NN) is a structure which is composed of multiple layers of neurons. Input is given to the first layer and last layer will give the output and one or more hidden middle layers are present which classifies the input more precisely and gives more accurate output. Support Vector Machine (SVM), a statistical learning approach is used to classify heterogeneous data. The SVM will separate samples into classes based on labeled data in a multidimensional space.

Decision trees are constructed using discrete attributes as nodes for each branch of the tree. For continuous attributes, two branches are created based on the selected threshold value. Memory based learning can be case-based or instance-based. The metrics used to differentiate among them are distance, number of attributes selected, the implicit or explicit knowledge discovered and the voting scheme for memory representation and retrieval algorithms. Each branch is called recursively for the training examples and stops when all nodes belong to same class.

Machine Learning techniques easily identify the patterns in the input and make the learning process easy. Since learning process is continuous, there is no need of manual intervention and there will be a continuous improvement in the result. Machine Learning handles dynamic, multi-dimensional, multi-variety data under uncertain environments.

3. Related Work

In the past few decades, more interesting and significant research has been carried out in the field of WWTP. This section discusses some of the works carried out in WWTP using Machine Learning techniques.

3.1 Neural Networks (NN)

Dario Torregrossa et al. [1] present a method on Neural Networks. The approaches followed for data-based cost modeling are linear, exponential, and logarithmic. The relationship between wastewater inflow and contaminant present are analyzed using a linear approach through least square regression method. The cost of constructing the facilities is analyzed. To consider uncertainty in data, linear regression approach is used. The cost function for WWTP construction, operation and maintenance is considered to account the impact. Exponential models investigate the impact of inaccurate plant design on the energy. Ageing factor is considered when modeling energy cost of WWTPs, investment and maintenance costs of water reuse technology are done using logarithmic form of cost function. Energy costs are based on model rather than data as per assessment. The decision-making process done to save energy supports dynamic model with energy costs. The food waste present in water analyses the sensitivity between input and output values. The method impacted cause and effect relationship and is used for control of cost usage using machine learning model. The challenging task is to efficiently analyze the relationship between parameters to establish a function. The regression models provide better results when the functions are not based on fixed functions. The cost parameter is suspected to impact the efficiency more.

Neural Networks (NN) and Random Forest (RF) are used as regression and classification algorithms. Random Forest is constructed using several decision trees, where input parameters are used to insert internal nodes. Selection of input parameters is done randomly to generate decision trees. Mean is calculated as a result. The algorithm splits dataset into training and testing sets, and then coefficient of determination (R2) is calculated for test set. The number of neurons(N) and R2 is stored and the steps are

repeated for 100 iterations. If R2 >0.75, model is stored. For RF, R2=0.82 and for NN, R2=0.8 which shows that both methods have similar impact on the input parameters.
Saif Al Aania et al. [2] introduces Artificial Neural Networks (ANN) which simplifies process of desalination of water. The input parameters for ANN are dissolved organic carbon, temperature, pH, and bromide concentration. These parameters are associated with weight and bias. In each epoch output is optimized. ANN predicts the salt and mixtures present in water. The properties of membrane are predicted under different circumstances and observed that it impacts performance. The model provided better result for both short-term and long-term procedure.

3.2 Remote Sensing and Machine learning

Shyamal S. Virnodkar et al. [3] presents a concept on Remote Sensing (RS) which is minimal cost and labor- intensive technique. Infrared thermometry method is proposed to assess plant water stress scale. Crop water stress detection methods are explained based on field measurements. Field measurement methods like soil moisture measurements are based on plants response in the form of growth. The meteorological variables are used to analyze water loss for a plant. Remote sensing based methods like Spectral indices based methods, Infrared thermometry, and Crop Water Stress Index based methods, LST based CWSI, Evapotranspiration based methods are non-destructive and not time intensive. A large amount of data is collected by sensors when monitoring the water level. To manage the data, Machine learning methods like SVM, Random forest classifier, eXtreme Gradient Boosting (XGBoost), Rotation forest, Oblique random forest, and Artificial neural network are used to manage large amount of data collected by sensors. The focus of SVM learner is on generating an optimal hyperplane and measuring the canopy temperature to measure its impact on water stress level.

Xusong Qin et al. [4] proposes wastewater quality monitoring system. The quality of water is measured in terms of turbidity. Turbidity test focuses on the transparency of water sample. The degree of transparency reduces when suspended particles droplets of oil, greases present in water are more. The Boosting-Iterative Predictor Weighting-Partial Least Squares (Boosting-IPW-PLS) method incorporated for turbidity measurement uses IPW for boosting scheme to remove the irrelevant variables by assigning small weights and building weighted models. After feature extraction is done, wastewater is treated under UV/Vis spectroscopy. The pollutants in water refracts describing the impact of input parameters.

3.3 Support Vector Machine (SVM)

Jun Inoue et al. [5] proposes supervised machine learning technique like Support Vector Machine (SVM) to detect anomaly for a Cyber-Physical System (CPS). The dataset is from Secure Water Treatment (SWAT) testbed. Sliding window method is applied in SVM. Based on log feature, each data in log entry has normal/abnormal label. Input to SVM has three parameters, size of the sliding window, weight controlling trade-off between wrong classifications and co-efficient of kernel. The detectors are trained first using log generated by SWAT. The training data is fed into one class SVM. Every new test data is labeled in the same way as training data. The logarithmic grid constructed primarily shows that normalization has huge impact on performance. The experiment is limited to the data from SWAT testbed and manually created samples of anomalies.

Hong Guo et al. [6] proposes SVM and ANN methods for prediction of effluent concentration in a wastewater treatment plant. A multilayer perceptron network is a non-linear algebraic function. A feed forward network is built using input variables like month, chemical oxygen demand, volumetric flow rate of inflow, suspended solid, pH,

total Nitrogen (T-N) concentration, temperature T-N of inflow. The input is normalized and ANN transfer function and SVM kernel function is applied. The modeled data is compared and measured to predict T-N concentration. As a result, ANN model is more reliable than SVM for decision making models.

Antonia Hadjimichael et al. [7] studies on Urban Water System (UWS). In Urban Water System, water is extracted from reservoirs and is treated to reuse. Decision support systems (DSS) are used for handling complex water treatment techniques. Sustainable development is adopted and UWS needs a measure to update economic standards, social and environmental sustainability. Technologies are used for conservation of water resources, reuse of wastewater and reduction in water consumption.

3.4 K-Nearest Neighbors (KNN)

Kejiang Zhang et al. [8] proposes K-Nearest Neighbors technique to predict the coagulant dosage in water treatment plants. Two types of SVM's regression are used. e-SVR and v-SVR uses radial basis function (linear) and polynomial function (sigmoid) respectively as kernel functions. KNN is used to predict small, medium, and large sized particles in wastewater treatment plants. The implementation procedure involved, identification of input, data analysis, normalization, differentiating training and testing dataset, model application. In KNN, the distance is used as metric. The prediction is done based on the distance between the training points. Performance of application of functions SVM and KNN for WTP's are almost same where sigmoidal function performs better for medium and large particles and for two small-sized particles.

3.5 Reinforcement Learning (RL)

Felix Hernandez-del-Olmo et al. [9] proposes a model-free reinforcement learning. An agent is free to set the dissolved oxygen point in water by interacting with the wastewater treatment plant and continuous learning. Agent controls oxygen by interacting with its environment. RL uses Markov decision process to model the environment. An agent resolves optimal policy by dynamic programming. Agent makes decision based on its observation and perception at that state and chooses an action to execute it. Model free RL-agent learns to behave on trial and error method.

3.6 Decision Trees

Joaquim Comas et al. [10] uses decision trees and memory-based learning. Decision trees are constructed based on the rule's induction. The decision tree included 243 examples and 24 input attributes. Among 24, 10 quantitative and 14 qualitative attributes. The average predictive accuracy obtained is 63.51% for the branch with large classes. In Memory based learning, 19 input attributes are considered. A 10-fold cross-validation of data is carried out by means of inductive classification.

5. Conclusion

An extensive review is carried out on application of various Machine Learning techniques in Wastewater Treatment Plant. Papers on the techniques such as Artificial Neural Networks, Random Forest, Reinforcement Learning, Support Vector Machine, K-Nearest Neighbors are analyzed and reviewed. All the methods applied provided better performance results. The input parameters considered for the experiments are different in each analysis. Cost of the energy consumption as an input parameter has less impact on the WWTP. The techniques used provided good comparison on the relationship between

the input parameters. Few techniques resulted in same performance outcomes. The survey done, covers application of various Machine Learning models in WWTP.

REFERENCES

[1] *Dario Torregrossaa, Ulrich Leopold, Francesc Hernández-Sancho, Joachim Hansen, "Machine learning for energy cost modelling in wastewater treatment plants", Elsevier, Journal of Environmental Management 223, (2018), pp. 1061-1067.*

[2] *Saif Al Aani, Talal Bonny, Shadi W. Hasan, Nidal Hilald, "Can machine language and artificial intelligence revolutionize process automation for water treatment and desalination?", Elsevier, Desalination Volume 458, (2017), pp. 84-96.*

[3] *Shyamal S. Virnodkar, Vinod K. Pachghare, V. C. Patil, Sunil Kumar Jha, "Remote sensing and machine learning for crop water stress determination various crops: a critical review", Springer Science+Business Media, LLC, part of Springer Nature, (2018), pp. 218-230.*

[4] *Xusong Qin, Furong Gao, Guohua Chen, "Wastewater quality monitoring system using sensor fusion and machine learning techniques", Elsevier, Water Research 46, (2016), pp. 1133—1144.*

[5] *Jun Inoue, Yoriyuki Yamagata, Yuqi Cheny, Christopher M. Poskitty, Jun Sun, "Anomaly Detection for a Water Treatment System Using Unsupervised Machine Learning", IEEE International Conference on Data Mining Workshops (ICDMW), (2017)*

[6] *Hong Guo, Kwanho Jeong, Jiyeon Lim, Jeongwon Jo, Young Mo Kim, Jong-pyo Park, Joon Ha Kim, Kyung Hwa Cho, "Prediction of effluent concentration in a wastewater treatment plant using machine learning models", Elsevier Journal of Environmental Sciences 32,(2015), pp. 90—101.*

[7] *Antonia Hadjimichael, Joaquim Comas, Lluís Corominа, "Do machine learning methods used in data mining enhance the potential of decision support systems? A review for the urban water sector", AI Communications 29, (2016), pp. 747—756.*

[8] *Kejiang Zhang, Gopal Achari, Hua Li, Amin Zargar, Rehan Sadiq, "Machine learning approaches to predict coagulant dosage in water treatment plants", Internal Journal System Assurance Engineering Management, 4(2), (2016), pp. 205—214.*

[9] *Felix Hernandez-del-Olmo, Elena Gaudioso, and Antonio Nevado, "Autonomous Adaptive and Active Tuning Up of the Dissolved Oxygen Setpoint in a Wastewater Treatment Plant Using", IEEE Transactions on Systems, Man and Cybernetics-Part C: Allications and Reviews, Vol. 42, No. 5, (2012), pp. 768—774.*

[10] *Joaquim Comas, Saso Dzeroski, Karina Gibert, Ignasi R.-Roda and Miquel Sànchez-Marrè "Knowledge discovery by means of inductive methods in wastewater treatment plant data", AI Communications, Vol 14, No. 1, 17(1), (2001), pp. 45—62.*