



Soft Computing Model Development using the DSN-M algorithm and PSO for the effluent COD of an anaerobic filter

Vijayabhanu Rajagopal¹, Radha.V²

¹ Assistant Professor, Department of Computer science, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore - 641043, India affiliation

² Professor, Department of Computer science, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore - 641043, India

Abstract: The building of highly complex computer simulation model for wastewater treatment systems is essential as the wastewater treatment processes are inherently dynamic. An accurate description of such processes involving large variations depends on the degree to which the wastewater must be purified. In this paper, back propagation neural network has been built to describe the anaerobic digestion process treating the cheese dairy wastewaters for the prediction of Chemical Oxygen Demand(COD). COD test is one of the methods of determining the organic pollutants in the wastewaters. The removal efficiency of COD estimates the performance of the model. The entire process in an anaerobic digestion for the prediction of COD includes data collection and pre-processing, feature selection and prediction of COD using Back Propagation Neural Network (BPN). During the pre-processing phase z-score normalization technique is used. In this paper, Dynamic Score Normalization technique with Mahalanobis distance (DSN-M) is proposed and it is compared with z-score normalization technique. On comparison, DSN-M achieved better results for normalization. To process further, the normalized data is analyzed by PSO algorithm for selecting the important features. The parameters of anaerobic digestion are optimized by Particle Swarm Optimization (PSO) for building an accurate model. The prediction of the anaerobic digestion effluent parameters viz., COD, with a good degree of accuracy is then obtained when trained using BPN. The degree of accuracy is judged for z-score and the proposed DSN-M with PSO. The prediction accuracy for z-score with PSO and the proposed DSN-M with PSO when trained using BPN showed 85.20% and 87.10% respectively. It is clear from the experimental results that the proposed DSN-M with PSO produced more satisfactory results than z-score.

Keywords: Anaerobic digestion, Chemical Oxygen Demand (COD), Back Propagation Neural Network (BPN), Dynamic Score Normalization with Mahalanobis distance (DSN-M), Particle Swarm Optimization component

I. Introduction

It is essential to treat the wastewater to remove the impurities before it is released, in order that it does not cause further pollution of water sources. Wastewater comes from several sources such as industries, hospitals, residences, etc., [1]. Water from agricultural and industrial outlets is the major source of wastewater. Some wastewaters are extremely complicated to purify than others; for instance, industrial wastewater treatment is a very complex process, at the same time domestic wastewater is comparatively simple to treat (although domestic waste has become more complex to purify, as a result of the increased amounts of pharmaceuticals and personal care products [2].

There are several levels of wastewater treatment; these are primary, secondary and tertiary levels of treatment. Majority of the municipal wastewater treatment services use primary and secondary levels of treatment and in certain cases they may use tertiary level treatments. The type and level of treatment possibly differ from one treatment plant to another [3].

Anaerobic digestion of wastewater is extremely suitable for industries releasing exceedingly concentrated (more than 1,500 mg COD/l) wastewaters, with nitrogen concentrations that are not excessively high [4]. The food and food processing industries, beer breweries, soft drink producing factories and paper producing or processing factories and certain chemical industries discharge wastewaters of this category. In this paper, cheese-dairy wastewater is taken for the treatment.

The disposal of large quantity of wastewaters without adequate treatment results in significant environmental pollution. The increased concerned towards environmental issues has motivated researchers to concentrate on the accurate operation and control of wastewater treatment systems. On the other hand, serious environmental and public health difficulties may result from improper operation of a wastewater treatment system, as discharging contaminated effluents to a receiving water body can cause or spread various diseases to human beings. This system involves several complex physical, biological and chemical processes. Often these

processes exhibit non-linear behaviors which are difficult to describe by linear mathematical models [5]. Additionally, the variability of the influent characteristics, in terms of composition, strength and flow rates, might influence model parameters and consequently operational control, significantly [6].

It is to be noted that in any water system, micro-organisms will use any organic and inorganic matter in the water and will generate biomass by utilizing oxygen present in the water. The oxygen necessary for the degradation of the organic matter biologically is called the Biochemical Oxygen Demand (BOD). The industrial and municipal wastewater effluents will possibly include several huge amounts of organic matter in terms of BOD and if discharged into natural water bodies, it can cause complete depletion of dissolved oxygen leading to the mortality of aquatic organisms.

The quantity of oxygen required to consume the organic and inorganic materials is called the Chemical Oxygen Demand (COD). There is a specific relationship between the COD and BOD under certain conditions and by determining the COD, the information regarding the BOD of the water/wastewater can be obtained.

COD is an essential test for evaluating the quality of effluents and wastewaters prior to discharge. COD test predicts the oxygen requirement of the effluents and is exploited for checking and control of discharges and for evaluating the performance of treatment plant.

In this paper, a model is built for the prediction of COD that describes the performance of anaerobic digester. The objective of this paper is to provide a proposed method viz., DSN-M normalization for pre-processing the dataset and compared with z-score normalization. The data set prepared using DSN-M normalization enhances the learning capability of the network with minimum error. The features in the multidimensional space of the anaerobic digester are optimized using PSO algorithm. The purpose is to train the BPN and to evaluate the model performance with regard to Mean Square Error(MSE) and regression coefficient. The minimum MSE evaluates the accuracy and the regression coefficient value, when obtained as 1, the difference between the actual and predicted values as minimum. The Daily records from the operation of anaerobic filter treating agro food waste water during a period of 6 months with a total number of 197 samples are taken for the study.

II. Methods and Materials

A. Artificial neural networks

Artificial neural networks (ANN), as powerful modeling methods have found a great deal of interest among the researchers in the last two decades. They can solve a wide range of problems and differential equations, particularly when the conventional approaches fail.

The data processing can extend over multiple layers and the final error criteria such as mean squared error (MSE) are calculated at the output layer which measures the accuracy of modeling [7].

A well trained network learns from the pre-seen experimental dataset (training data) and generalizes this learning beyond unseen data which is called 'prediction' [8]. Furthermore, ANNs are able to model non-linear behaviours and complex processes. This is highly important in anaerobic treatment and bio-processes due to the special hydrodynamics and non-linear nature of the anaerobic digestion [9]. Based on the mechanism of human nervous systems, the ANN models can be classified into three major groups; 1) feed-forward network, 2) recurrent network and 3) unsupervised network.

A new optical technique of testing the COD of wastewater with multi-wavelength at UV band was developed by YanghuaLv *et al.*, [10]. The normal UV254 depends on the UV light absorption by wastewater at 254nm based on the Beer-Lambert law. However not all of the wastewater has apparently absorption at 254nm. This technique uses back-propagation ANN to find out COD of the samples. Following the BP network is training by a sequence of wastewater samples which have comparable ingredients but different COD, the sample's absorbance at UV band and the particular wavelength which is associative with sample's COD is picked out [11]. These wavelengths can be utilized to compute the COD value and the result is almost nearer to standard chemical method.

A soft-sensing technique was developed by Zhou *et al.*, [12] using ANN with the purpose of detecting the vital parameters like Dissolved Oxygen (DO), COD and BOD. A BP neural network is trained using the testing data from a convenient treatment process. This system can be employed for real-time control of wastewater treatment.

Soft-sensing method based on ANN is proposed by Wang Wan-liang and Ren Min [13] to detect the quality parameters in on-line. The wastewater treatment technique is investigated systematically. ORP, DO, PH and MLSS which can be detected on-line are considered as the secondary variables. BOD, COD, N and P which cannot be identified on-line are considered as the primary variables. The BP neural network for soft-sensing is trained using the testing data. As a result, the system can be transformed into the alterations in the environment and employ the real time control of wastewater treatment. In this paper, the Backpropagation neural network is built for the prediction of COD.

B. Cheese-dairy wastewater

The cheese dairy wastewater used in this study was obtained from a middle-size cheese factory without whey recovery. It was then diluted with water to obtain the COD concentration in the range of 30 g/L and particulate COD (COD_p) of about 3-4 g/L. pH of the feed was adjusted to 6-6.5. Whey contains lactose (5%), proteins (1%), lipids (0.2%), ash (0.6%) and water (94%), along with most of the minerals present in milk and water soluble vitamins [14].

C. Experimental set-up

Anaerobic filter (AF) was used as a high rate anaerobic system to treat the cheese dairy wastewater. AF is a bioprocess in which the bacteria colonize in and around certain materials, which increase the usable surface area for the bacterial growth [15]. This kind of reactor allows a higher volume load and is suitable for wastewater treatment of high organic material content. The laboratory-scale AF with 10 L effective volume was operated at mesophilic conditions (33°C) at increasing organic loading rates. About 80% of the reactor volume was filled with a cylindrical polyethylene packing media (30 mm diameter). This reactor was operated for a total period of more than 6 months. Based on the COD removal, the OLR was increased from 0.4 to a maximum of 17 g COD/L.d. A soluble COD removal of 80% was considered as the threshold limit for the successful operation of the reactor.

III. Model Development

In this paper anaerobic digestion of cheese-dairy wastewater was considered as a representative agro-food industrial effluents and the several steps involved in the model development are shown in figure 1.

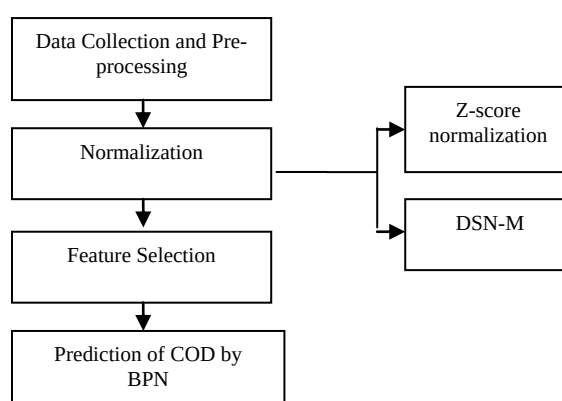


Figure. 1 Steps in the development of a model

A. Data Collection and Pre-processing

Daily records from the operation of an anaerobic filter treating cheese whey agro-food wastewater during a period of 6 months were obtained [15]. Processing the data is an essential step to establish the validity and quality consisting of variable measurements. Meaningless data are the results of the AF due to endless sources of disturbances. The sources may consist of electromagnetic interference, defective installation, insufficient maintenance or erroneous use and handling of the measuring system. Therefore, before any analysis of the measurements, it is necessary to preprocess the raw data.

Data preprocessing includes data cleaning, normalization, transformation, feature extraction and selection, etc. If data is irrelevant, redundant, noisy or unreliable, then learning of the relationships between the parameters during the training phase are more difficult. Training ANN until the error converges to a minimum does not mean that the NN will respond positively to the application of a test data set. Data has to be pre-processed before it is inputted to a neural network.

B. Missing values and outliers

If there are relatively few missing points, there are some methods to complete the series by filling the incomplete dataset with mean or median.

Regression will work best if the number of WWTS parameters having missing values in their records is small. Outliers can sometimes be accommodated in the data through the use of trimmed means. In this section, brief outline of the K-Means clustering is given. The results of applying the algorithm is the removal of the missing values and outliers and then clustering is used to group the related data for processing [16].

The approach adopted in this paper to replace the missing values is the interpolation method.

a. K-Means clustering

K-Means methodology is a commonly used clustering technique. In this analysis the user starts with a collection of samples and attempts to group them into 'k' Number of Clusters based on certain specific distance measurements [17]. The general objective is to obtain that partition which, for a fixed number of clusters, minimizes the square-error. K-Means is a commonly used partition algorithm with square-error criterion, which minimizes the following error function.

$$J_m = \sum_{i=1}^N \sum_{j=1}^C u_{ij} m \|x_i - c_j\|^2, 1 \leq m < \infty \quad (1)$$

where $\|x_i - c_j\|^2$ is a chosen distance measure between a data point x_i and the cluster center c_j , is an indicator of the distance of the n data points from their respective cluster centres. K-means clustering is used in this paper to focus on the removal of the outliers.

C. Z-Score Normalization

The most frequently used score normalization technique is the z-score that is computed using the arithmetic mean and standard deviation of the particular data. This technique can be estimated to execute well if previous knowledge regarding the average score and the score variations of the matcher is available. If there is no previous knowledge regarding the nature of the matching algorithm, then it is necessary to calculate approximately the mean and standard deviation of the scores. The normalized scores are given by

$$s'_k = \frac{s_k - \mu}{\sigma} \quad (2)$$

Where μ represents the arithmetic mean and σ represents the standard deviation of the given data. On the other hand, both mean and standard deviation are responsive to outliers and as a result, this technique is not robust. Z-score normalization does not assure a general numerical range for the normalized scores of the different matchers. In case of an arbitrary distribution, mean and standard deviation are adequate estimates of location and scale, correspondingly, but are not finest.

D. Proposed Dynamic Score Normalization with Mahalanobis Distance (DSN-M)

The proposed DSN-M technique is discussed in this section. DSN-M uses Dynamic Time Warping (DTW) to equal the cumulative recognition rate of a classifier against a standard curve that is identical for all the classifiers [18]. DSN-M considers the accumulated Gaussian probability function as standard curve ($\mu = 0, \sigma = 1$). Now, consider

$$R = (r_1, r_2, \dots, r_L) \quad (3)$$

is the L-point discrete version of the accumulated recognition function and

$$G = (g_1, g_2, \dots, g_L) \quad (4)$$

is the L-point discrete version of the accumulated probability function of the standard Gaussian distribution. A warping function between R and G is any series of couples of indexes recognizing points of R and G to be joined [19].

$$W(R, G) = c_1, c_2, \dots, c_T \quad (5)$$

in which $c_t = (i_t, j_t)$, being T, i_t, j_t integers, $1 \leq t \leq T, 1 \leq i_t \leq L, 1 \leq j_t \leq L$.

Now, if a distance measure $d(\cdot, \cdot)$ is considered between points of R and G ,

$$d(c_t) = d(r_{i_t}, g_{j_t}) \quad (6)$$

$d(c_t)$ can be combined to $W(R, G)$ the dissimilarity measure

$$D_{W(R, G)} = \sum_{t=1}^T d(c_t) \quad (7)$$

DSN-M uses Mahalanobis distance. The elastic matching procedure identifies the optimal warping function $W^*(R, G) = c_1^*, c_2^*, \dots, c_{T^*}^*$ which satisfies the conditions of

- i. monotonicity,
- ii. continuity,
- iii. boundary conditions,

and for which it results:

$$D_{W^*(R,G)} = \min_{W(R,G)} D_{W(R,G)} \quad (8)$$

Now consider $W^*(R, G)$ be the optimal warping function for the matching between R and G , given a matching score a_i , $i = 1, 2, \dots, n$, its normalized value is given as:

$$a'_i = \frac{1}{\text{card}(G_i)} \cdot \sum_{g_q \in G_i} g_q \quad (9)$$

Where G_i is the collection of samples of G related with the sample r_i of the accumulated recognition function, i.e.:

$$G_i = \{g_q | (q, i) \in W^*(R, G)\} \quad (10)$$

The normalized value of a score a_i is specified by the average value of the points of the accumulated Gaussian function which is equal to the sample r_i , being r_i the sample of the accumulated recognition function that corresponds to a_i . As a result, DTW is exploited to map the value of the accumulated recognition function r_i on the standard curve. The average of the values on the standard curve that corresponds to r_i is taken as the normalized value a'_i corresponding to a_i .

E. Feature subset selection using PSO

Machine learning involves optimization of data, thus producing acceptable prediction accuracy. Feature selection reduces the number of features and also eliminates irrelevant noisy data. Thus the optimal data subset of features among the total features is found rendering the best performance. By feature selection, the neural network with the improved learning performance enhances the prediction performance when compared to the prediction

In this work, Particle Swarm Optimization (PSO) is used to implement a feature selection. PSO is an evolutionary computation technique developed by Kennedy and Eberhart in 1995 [20]. Initial simulations were changed to develop the original version of PSO. In order to produce the standard PSO, Shi introduced inertia weight into the particle swarm optimizer [21]. A population of random solutions is initialized to PSO called 'particles'. Each particle can be treated as a point in a S-dimensional space. The i th particle is represented as $X_i = (x_{i1}, x_{i2}, \dots, x_{is})$. The best previous position (pbest, the position giving the best fitness value) of any particle is recorded and represented as $P_i = (p_{i1}, p_{i2}, \dots, p_{is})$. The index of the best particle among all the particles in the population is represented by the symbol 'gbest'. The rate of the position change (velocity) for particle i is represented as $V_i = (v_{i1}, v_{i2}, \dots, v_{is})$. The particles are manipulated according to the following equation:

$$v_{id} = w * v_{id} + c_1 * \text{rand}() * (p_{id} - x_{id}) + c_2 * \text{rand}() * (p_{gd} - x_{id}) \quad (11)$$

$$x_{id} = x_{id} + v_{id} \quad (12)$$

where $d = 1, 2, \dots, S$, w is the inertia weight which is a positive linear function of time changing according to the generation iteration. Appropriate selection of the inertia weight provides a balance between global and local exploration. This results in less iteration on average to measure the most favorable solution. The acceleration constants c_1 and c_2 in equation (11) indicates the weighting of the stochastic acceleration terms which pull each particle towards the pbest and gbest regions. High values result in abrupt movement towards target regions while low values allow particles to roam away from the target regions before being pulled back. $\text{rand}()$ is a random function which lies in the range $[0, 1]$.

Particles' velocities on each dimension are limited to a maximum velocity, $V_{\max}$. [22]. It determines how large steps through the solution space in each particle are allowed to take. If $V_{\max}$ is too high, particles might rush towards good solutions. On the other hand, If $V_{\max}$ is too small, particles may not explore adequately beyond locally good regions. It could be trapped in local optima. The first part of equation (11) provides the "flying particles" with a degree of memory capability allowing the exploration of new search space areas. "cognition" part is the second part, which represents the private thinking of the particle itself. Final part is the "social" part, which indicates the collaboration among the particles. Equation (11) is used to calculate the particle's new velocity according to its previous velocity and the distances of its current position from its own best experience (position) and the group's best experience. According to equation (12), the particle flies toward a new position. The performance of each particle is measured according to a pre-defined fitness function.

PSO Algorithm:

Inputs:

m : the swarm size; c_1, c_2 : positive acceleration constants; w : inertia weight

MaxV : maximum velocity of particles

MaxGen : maximum generation

MaxFit : maximum fitness value

Output:

```

Pgbest: Global best position
Algorithm:
Begin
Swarms  $\{x_{id}, v_{id}\}$  =Generate(m); /* Initialize a population of particles with random
positions and velocities on S dimensions*/
Pbest(i)=0; i= 1,...,m, d = 1,...,S
Gbest=0; Iter=0;
While(Iter<MaxGen and Gbest<MaxFit)
{ For(every particle i)
{ Fitness(i)=Evaluate(i);
IF(Fitness(i)>Pbest(i))
{Pbest(i)=Fitness(i); pid= xid; d = 1,...,S }
IF(Fitness(i)>Gbest)
{Gbest=Fitness(i); gbest=i;}
}
For(every particle i)
{ For(every d ){
 $v_{id} = w * v_{id} + c_1 * \text{rand}()$ 
 $* (p_{id} - x_{id}) + c_2 * \text{rand}()$ 
 $* (p_{gd} - x_{id});$ 
IF( $v_{id} > \text{MaxV}$ ) {  $v_{id} = \text{MaxV}$ ;}
IF( $v_{id} < - \text{MaxV}$ ) {  $v_{id} = - \text{MaxV}$ ;}
 $x_{id} = x_{id} + v_{id}$ 
}
}
Iter=Iter+1;
}/*rand() and Rand() are two random functions in the range [0,1]*/
Return P_{gbest}

```

End

PSO is easy to implement and there are few parameters to adjust. Thus by using PSO, as a feature selection, reduces noise in the dataset to improve prediction accuracy.

F. Prediction of COD by BPN

Neural networks have been effectively applied to various biochemical processes [23]. They have a distinctive capability to model non-linear dynamic systems without having need of a structural knowledge of the process to be modeled. Neural networks can map a collection of input patterns onto an equivalent set of output patterns on the basis of historical data from any given system.

Two widely used networks for modeling the non-linear problems in engineering systems are the Back Propagation and Radial Basis Function (RBF) neural networks. Back propagation is the training technique usually used for this purpose. Levenberg–Marquardt algorithm is used to train the Back Propagation.

IV. Results and discussions

The data sets obtained from an anaerobic filter are high-dimension. The high-dimension data set, which means they consist of many variables and are often highly correlated. These high-correlations take a longer time to train and forecast effluent COD, and it may also introduce unwanted noise.

Having large number of input variables leads to complex design issues as well as longer time to train the data. The optimization of such data sets reduces any redundant attributes thus determining the proper inputs for the model. Hence, after pre-processing, n number of features were selected and given as input to back propagation neural network. When two features are highly dependent on each other, they are said to have high redundancy. Low redundant features are found in order to find mutually exclusive features. A mutually exclusive feature has low redundancy and high relevance when compared to other features and highly dependent on the target class. Such features are not removed from the feature subset.

Daily records from the operation of an anaerobic filter treating cheese dairy wastewater during a period of 6 months were obtained as shown in table 1.

TABLE 1: DATA SET FROM AN ANAEROBIC FILTER (SAMPLE DATA)

Operating conditions				Influent values				
flow rate	OLR	HRT	CODs	TSS	VSS	PCOD	CODt	pH
0.135	0.49	74.07	32.45	2.773	1.643	3.96	36.41	6.89
0.135	0.49	71.43	32.45	2.773	1.643	3.96	36.41	6.82
0.140	0.51	85.47	32.45	2.773	1.643	3.96	36.41	6.78
0.117	0.43	74.63	29.50	2.506	1.485	3.40	32.90	6.76
0.134	0.44	80.65	24.42	2.269	1.344	5.37	29.79	6.80
0.124	0.37	80.65	24.42	2.269	1.344	5.37	29.79	6.80
0.124	0.37	80.65	24.42	2.269	1.344	5.37	29.79	6.80
0.124	0.37	57.80	29.28	2.598	1.539	4.83	34.11	7.02
0.173	0.59	100.00	29.28	2.598	1.539	4.83	34.11	7.06
0.100	0.34	83.33	29.28	2.598	1.539	4.83	34.11	7.05
0.120	0.41	83.33	24.80	2.216	1.313	4.30	29.10	7.04
0.120	0.35	91.74	24.80	2.216	1.313	4.30	29.10	7.05
0.109	0.32	91.74	24.80	2.216	1.313	4.30	29.10	7.05
0.109	0.32	91.74	24.80	2.216	1.313	4.30	29.10	7.05
0.109	0.32	52.91	28.68	2.292	1.358	1.41	30.09	7.04
0.189	0.57	46.95	28.68	2.292	1.358	1.41	30.09	6.98
0.213	0.64	47.62	28.68	2.292	1.358	1.41	30.09	6.92
0.210	0.63	45.45	30.39	2.463	1.459	1.95	32.34	6.96
0.220	0.71	39.68	24.20	2.209	1.309	4.80	29.00	6.96

The model was trained with the data in the training subset with influent COD levels, flow rate and OLR to optimize the network weights so as to minimize the appropriate error function.

The model was developed by using Matlab software and the performance of Back propagation model for z-score and DSN-M normalization techniques was evaluated by calculating the MSE, Regression coefficient and Average prediction accuracy and the features selected.

It is observed from the Table 2 that DSN-M normalization technique performs well than z-score based on MSE, regression coefficient, average prediction accuracy and the selection of number of features. Thus the learning performance is improved by DSN-M technique.

TABLE 2: COMPARISON OF Z-SCORE WITH PSO AND DSN-M WITH PSO WITH VARIOUS PARAMETERS

Methodology	Training time (sec)	Testing time (sec)	Accuracy (%)	No. of features	Features selected	MSE	Regression coefficient
Z-Score with PSO	8.912	0.803	85.20	10	8	46.86	0.72
DSN-M with PSO	5.811	0.388	87.10	10	6	43.76	0.81

A. DSN-M with PSO

The dataset normalized with DSN-M consists of all the features involved in the anaerobic digester. The multidimensional dataset when trained using BPN degrades the learning performance of the network without the proper selection of the required features. For this purpose, PSO, an optimization technique is applied for dimensionality reduction of the dataset. The Computational overhead is greatly reduced. The backward elimination search strategy is used to find the best feature subset. The candidate feature subset found is then evaluated using BPN. It is clear from the observations that the training time required for BPN after normalizing with DSN-M and performing the dimensionality reduction using PSO is found to be 5.811 (secs) and the testing time is 0.388 (secs). The total number of features was 10 out of which PSO found 6 best feature subsets for the better performance of the network. The MSE and regression coefficient values of the network with PSO are found to be 43.76 and 0.81 respectively.

V. Conclusion

Water is one of the useful natural resources and it is facing several disputes with the fast development of large number of industries. Almost all the industries discharge the wastewater into nearby water bodies without treating it. This is turned out to be a biggest threat to human lives and many other aquatic organisms. Hence it is necessary to treat or purify the water before discharging it into other water bodies. In this paper, back propagation neural network has been built to predict the COD. COD is an essential test for evaluating the quality

of effluents and wastewaters prior to discharge. In this paper, proposed Dynamic Score Normalization technique with Mahalanobis distance (DSN-M) and it is compared with z-score normalization technique. This paper investigates the effect of proposed DSN-M technique on the performance of a prediction level of COD of BPN. The selection of features from the entire data is then continued using PSO and the learning overhead is reduced when trained using BPN. Both continuous and discrete variable are considered and it is computationally inexpensive, since memory and CPU speed requirements are low. The average prediction accuracy for z-score with PSO and the proposed DSN- with PSO when trained using BPN showed 85.20% and 87.10 % respectively. The BPN with DSN-M normalized data and the best feature subset selected by PSO predicted the COD with satisfactory results on consideration of MSE, regression coefficient, prediction accuracy and the selection of number of features.

References

- [1] Hamed.M.M, Khalafallah. M.G. and Hassanien.E.A(2004), "Prediction of wastewater treatment plant performance using artificial neural networks", Environmental Modelling and Software, Vol. 19, No. 10, Pp. 919-928.
- [2] Metcalf.E (1996), "Wastewater Engineering, Treatment, Disposal and Reuse", 5th Edition, McGraw Hill, NY.
- [3] Sawyer. C.N., McCarty. P.L. and Parkin. G.F.(1994), "Chemistry for Environmental Engineering", 4th Edition, McGraw-Hill International Editions.
- [4] Sarah Volkman(2003), "Sustainable Wastewater Treatment and Reuse in Urban Areas of the Developing World", Field Engineering in the Developing World, Michigan Technological University, Master's International Program, Pp. 1-17.
- [5] Plazl. L., Pipus. G., Grolka. M. and Koloini.T.(2003), "Parametric sensitivity and evaluation of a dynamic model for single-stage wastewater treatment plant", Biotechnology for the Environment: Wastewater treatment and Modelling, Waste gas handling, Pp. 65-72.
- [6] Hamoda.M.F., Al-Gusain.I.A. and Hassan. A.H.(1999), "Integrated wastewater treatment plant performance evaluation using artificial neural network", Water Science and Technology, Vol. 40, Pp. 55-69.
- [7] Hertz, J., Krogh, A., Palmer, R. G.(1991) Introduction to the Theory of Neural Computation. Addison-Wesley Publishing Company, Redwood City, NJ.
- [8] Haykin, S(1994) "Neural Networks: A Comprehensive Foundation", Mac- millan, New York.
- [9] Holubar, P., Zani, L., Hager, M., Froschl, W., Radak, Z. and Braun, R(2002). "Advanced controlling of anaerobic digestion by means of hierarchical neural networks" Water Research. 36: 2582–2588.
- [10] YanghuaLv, Zhen Xiang, Ke Chen and LejiShao(2011), "A new method of COD testing with BP neural network", International Conference on Electronics and Optoelectronics (ICEOE), Vol. 4, Pp. 272 - 275.
- [11] Jun Fang and Liankui Dai(2004), "Rapid detection of chemical oxygen demand using least square support vector machines", Fifth World Congress on Intelligent Control and Automation, Vol. 5, Pp. 3810–3813.
- [12] Zhou. L., Fang. Y, Xie. L. and Zhang. S. (2006), "A Soft-sensing Method Based on BP Neural Network for Improving Dissolved Oxygen Measurement", 1ST IEEE Conference on Industrial Electronics and Applications, Pp. 1–5.
- [13] Wang Wan-liang and Ren Min(2002), "Soft-sensing method for wastewater treatment based on BP neural network", Proceedings of the 4th World Congress on Intelligent Control and Automation, Vol. 3, Pp. 2330–2332.
- [14] Thanial. J.V., Torrijos. M., Habouzit.F. andMoletta. R. (2007), Treatment of distillery vinasse in a high rate anaerobic reactor using low density polyethylene supports, Water Sci. Tech., 56(2) 17–24.
- [15] Rajagopal.R (2008).Treatment of Agro-Food Industrial Wastewaters Using UAF and Hybrid UASB-UAF Reactors", PhD Thesis.Indian Institute of Technology Roorkee, Roorkee, India.
- [16] Barnett.V and Lewis.T (1994). Outliers in statistical data , 3rd Ed., Wiley, Chichester, U.K.
- [17] Rabee Rustum and Adebaye J. Adeboye(2007), Replacing Outliers and Missing Values from Activated Sludge Data using Kohonen Self-Organizing Map, ASCE,909 –916.
- [18] Modugno. R., Pirlo. G. andImpedovo. D.(2010), "Score normalization by Dynamic Time Warping", IEEE International Conference on Computational Intelligence for Measurement Systems and Applications (CIMS),Pp. 82-85.
- [19] Kennedy. J., Eberhart. R.C.(1995), "A new optimizer using particle swarm theory", Sixth International Symposium on Micro Machine and Human Science. Nagoya, pp. 39-43.
- [20] Kennedy. J., Eberhart. R.C.(1995), "Particle Swarm Optimization", Proc IEEE Int. Conf. On Neural Networks, Perth, pp. 1942-1948.
- [21] Eberhart. R.C., Shi. Y.(2001), "Particle swarm optimization: Developments, applications and resources", Proceedings of the IEEE International Conference On Evolutionary Computation. Seoul, pp. 81-86
- [22] Xiangyang Wang, Jie Yang, XiaolongTeng, Weijun Xia(2007), "Feature Selection based on Rough Sets and Particle Swarm Optimization", In Pattern Recognition Letters, Vol. 28, No. 4.
- [23] Lee. D.S. and Park. J.M.(1999), "Neural network modelling for on-line estimation of nutrient dynamics in a sequentially-operated batch reactor", J. Biotech., Vol. 75, Pp. 229–239.