Available online at www.sciencedirect.com

ScienceDirect

Journal of Hazardous Materials 152 (2008) 1268–1275

Journal of Hazardous Materials

www.elsevier.com/locate/jhazmat

Prediction of biosorption efficiency for the removal of copper(II) using artificial neural networks

N. Prakash*, S.A. Manikandan, L. Govindarajan, V. Vijayagopal

Department of Chemical Engineering, Annamalai University, Annamalai Nagar 608002, Tamilnadu, India

Received 18 December 2006; received in revised form 30 June 2007; accepted 2 August 2007
Available online 11 August 2007

Abstract

Various low-cost adsorbents have been used for removing Cu(II) ions from aqueous solutions for the treatment of copper containing wastewaters to remove organic compounds and color. Sawdust is an impressive adsorbent in terms of adsorption efficiency, cost and availability; hence the use of sawdust as biosorbent has been widely studied. Many earlier investigations tried to correlate the experimental data with available models or some modified empirical equations, but these results were unable to predict the values of parameters from a single equation. Artificial neural networks (ANN) are effective in modeling and simulation of highly non-liner multivariable relationships. A well-designed and very well trained network can converge even on multiple number of variables at a time without any complex modeling and empirical calculations. In this present work ANN is applied for the prediction of percentage adsorption efficiency for the removal of Cu(II) ions from aqueous solutions by sawdust. Artificial neural network model, based on multilayered partial recurrent back-propagation algorithm has been used. The performance of the network for predicting the sorption efficiency of sawdust for copper is found to be very impressive.

Keywords: Biosorption; ANN; Neural network; Simulation; Copper(II) (Cu(II))

1. Introduction

Biosorption of metals from aqueous solutions can be considered as a promising technology in industrial wastewater treatment. It is based on the ability of biological materials to accumulate metal ions from wastewater by either metabolically mediated or physicochemical pathways of uptake [1]. This innovative process uses biomaterials which are either abundant in nature or wastes coming from industrial production and biological processes such as fermentation and water treatment.

The advantage of biosorption is that it uses biomass and industrial plant waste which are cheap and abundant. Broad-range biosorbents can collect most of the metal ions from the solution, and a certain concentration of a specific metal could be achieved either during the adsorption uptake by manipulating the properties of a biosorbent, or upon desorption during the regeneration cycle of these biosorbent. There have been numerous studies on the adsorption of metal ions from aqueous solutions by microbial and plant biomass [1–6], capable of collecting toxic heavy metals (Cd^{2+}, Hg^{2+}, Zn^{2+}, Pb^{2+}), precious metals (Au^{3+}, Pd^{2+}, Ag^{+}) and base metals (Co^{2+}, Ni^{2+}, Cu^{2+}).

The mechanism of biosorption is highly complex and is difficult to model and simulate using conventional mathematical modeling. This is mainly due to the interaction of more number of adsorption process variables, and hence the resulting relationships are highly non-linear [11]. Adsorption isotherms are inadequate to accurately predict the extent of adsorption and reproduction of results.

Application of artificial neural networks (ANN) has been considered as a promising tool because of their simplicity towards simulation, prediction and modeling. The advantages of ANN are that they require less time for development than the traditional mathematical models, the need for extensive experimentation is avoided as limited numbers of experiments are sufficient to predict the degree of non-linearity and their ability to learn complex relationships without requiring the knowledge of the model structure [7]. ANN models could describe adsorption systems better than general rate models [10]. Even the adsorption isotherms can be represented by neural networks [12]. So, it is preferable to use a non-parametric technique such as a

* Corresponding author. Tel.: +91 9443240210.
E-mail address: busiprakash@hotmail.com (N. Prakash).

0304-3894/$ – see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.jhazmat.2007.08.015

back-propagation neural network to represent such an equilibrium relationship [11].

The present work investigates the implementation of neural networks for the prediction of biosorption efficiency for the removal of copper using an effective and cheap biosorbent namely sawdust. The network results were compared with those obtained through experiments.

2. Artificial neural networks

Artificial neural networks are known for their superior ability to learn and classify data. The inspiration of neural networks came from studies on the structure and function of the brain and nerve systems as well as the mechanism of learning and responding. The potential applications include prediction, classification, data association, data conceptualization, data filtering and optimization.

There are two broad classifications of neural network structures [13]:

(i) *External structure:* Describes the overall, or macro arrangement of connections between inputs, outputs and hidden layers that compose the output. The several general external arrangements are, single-input and single-output (SISO), multiple-input and single-output (MISO) and multiple-input and multiple-output (MIMO).

(ii) *Internal structure:* Refers to the actual connections between individual nodes both within and between layers. Any node can be connected with any node in the network. The relative position of the origin to the endpoint of the connection defines the network's internal structure. There are three types [13] of connections that are in use: inter-layer, intra-layer and recurrent:
   - Inter-layer connection: Outputs from nodes on one layer feed into nodes in a different layer can be further classified as feed forward and feed back.
   - Intra-layer connection: Outputs from nodes in one layer feed into nodes in that same layer.
   - Recurrent connection: Outputs from a node feed into itself as inputs.

Generally the complexity of the problem considered and interaction of the variables involved decides the selection of network structure and type of connection between layers and nodes. Further, the selection of optimum number of hidden layers and number of neurons in each hidden layer may be justified from the performance of the network and by comparing the resulting mean square and standard deviation errors.

The different types of networks [15] based on their incremental complexity are: feed-forward network, recurrent network, stochastic network, modular networks and a few special types – holographic associative memory, instantaneously trained, spiking networks, etc. – which do not fit in any of these categories.

As biosorption influenced by number of process variables which hold complex non-linear relationships among them so as to affect the sorption process, any simple feed-forward network may not be sufficient to handle the prediction efficiently [8]. A capable recurrent network with suitable training algorithm may perform the task better.

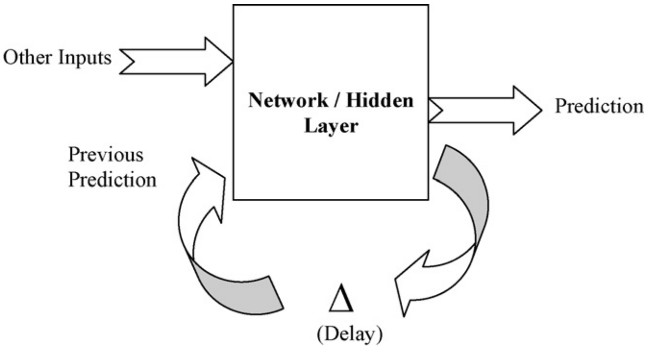

Fig. 1. Recurrent pattern.

2.1. Recurrent network

A recurrent network [14] is one where the recurrent usage of output occurs in such a way that the momentary output of the network or hidden layer is fed back as succeeding additional input along with other inputs. In a recurrent network, therefore, there will be at least one cyclic path (Fig. 1). Recurrent networks can be used to process sequences of data; sequences include samples of data at successive time intervals.

There are two types of recurrent networks:

- Fully recurrent: Here the output of the network is fed back in to the network (e.g., Jordan network)
- Partially recurrent: Here the recurrent connection is from the hidden layer to the state vector, not from the output layer (e.g., Elman network)

Fully recurrent networks feed back the hidden layer to it self. Partially recurrent networks start with a fully recurrent net and add a feed-forward connection that bypasses the recurrence, effectively treating the recurrent part as a state memory. These recurrent networks can have an infinite memory depth and thus, find relationships through time as well as through the instantaneous input space. Recurrent networks are the state-of-the-art in non-linear time series prediction, system identification and temporal pattern classification.

2.2. Elman network

Elman networks [9] generally follow partially recurrent multi-layer back-propagation, with the addition of a feedback connection from the output of the hidden layer to its inputs (Fig. 2). This feedback path allows Elman networks to learn, recognize and generate temporal patterns and spatial patterns.

The number of state vector units must equal the number of hidden units, state vector units look like input units to the back-propagation algorithm. At time zero, the current state units are set to a fixed value (0.5 for Elman), and the input units are set to the pattern representing the first sequence member. At subsequent time steps (Δ), the input units are set to the next sequence

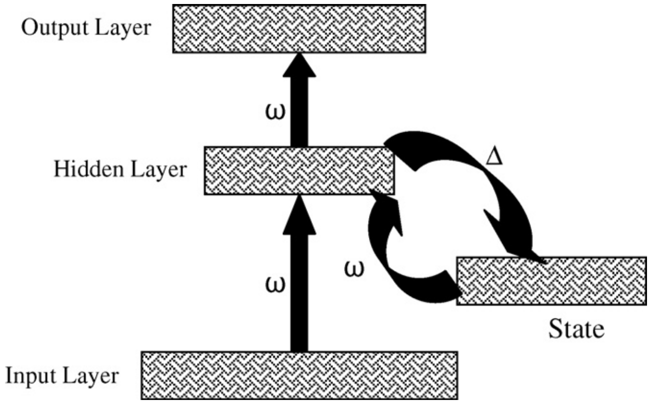

Fig. 2. Partial recurrent pattern.

member, and the current state units are set to the previous hidden unit values, back-propagation works as usual.

The Elman network utilizes specific implementation of transfer functions; it has Tanh neurons in its hidden (recurrent) layer and Linear Sigmoid neurons in its output layer. This combination is special in recurrent networks, these transfer functions can approximate any function (with a finite number of discontinuities) with arbitrary accuracy. The only requirement is that the hidden layers must have enough neurons. More hidden neurons are needed as the function being fit increases in complexity.

The Elman network differs from conventional networks in that the hidden layers have recurrent connections. The delay between these connections store values from the previous time step, which can be used in the current time step.

2.3. Back-propagation training of recurrent networks

In pure feed-forward networks the back-propagation [13] will be carried out in such a way that the network compares its output with the training data, then the network calculates the amount of error between its predicted output and the actual output. The network works backwards through the layers, adjusting the weight factors according to how much error it has calculated in its output. Once all of the weight factors have been adjusted, the network works in a forward path, taking the same input data to predict the output, based on the new weight factors. The network again calculates the error between the predicted and actual outputs. It adjusts the weight factors and the process continues, iteratively, until the mean square errors between the predicted and actual outputs have been minimized. In this case the network does not maintain a short-term memory, where as a recurrent network does.

In the original experiments presented by Elman [9] truncated back-propagation was used. This basically means that the delayed state value $y_j(t-1)$ was simply regarded as an additional input, any error at the state layer $\delta_j(t)$, was used to modify weights from this additional input slot.

Errors can be back-propagated even further. This is called back-propagation through time (BPTT), where all recurrent weights will be duplicated spatially for an arbitrary number of time steps (τ). Consequently, each node which sends activation along a recurrent connection has – at least – τ number of copies as well. Errors are thus back-propagated according to the fillowing equation:

$$\delta_{pj}(t-1)=\sum_{h}^{m}\delta_{ph}(t)u_{hj}f(y_{pj}(t-1))$$

where h is the index for the activation receiving node and j for the sending node (one time-step back).

It is important to note, however, that after error deltas have been calculated, weights are folded back adding up to one big change for each weight. Obviously there is a greater memory requirement and the case is larger the τ greater the memory will be. But still, in practice very large τ may lead to vanishing gradient effect. As such for each layer the error is back-propagated through the network, the error gets smaller and smaller until it diminishes completely.

2.4. Testing the network

An important aspect of developing neural networks is in the determination of the performance after the training is complete. The performance of the network is evaluated on the basis of two main criteria [7,13] based on:

(i) Recollection of the predicted output from the set of training data called as the recall step.
(ii) Ability of the network to predict output for the data not presented in the training set called as generalization step.

In the recall step, the network's performance is evaluated by recalling (retrieving) specific initial input used in training. Thus, a previously used input pattern was introduced to the trained network. The network then attempts to predict the corresponding output. Recall testing is so named because it measures how well the network can recall what it has learned. Generalization testing is conducted in the same manner as recall testing; however, now the network is given input data with which it was not trained. Generalization testing is so named because it measures how well the network can generalize what it has learned and form rules with which to make decisions about data it has not previously seen.

3. Materials and methods

Sawdust of *Mangifera indica* (mango tree) has been most impressive among the efficient bio-adsorbents so for reported. It was used effectively [6] as an adsorbent for the removal of Cu(II) from aqueous solution. Sawdust is washed to clean the adhering dirt, rinsed thoroughly with double-distilled water and finally heated in an air oven at 100–105 °C for 24 h. After drying, the adsorbents are sieved by using 100 mesh sieve.

3.1. Variables affecting the biosorption

The various factors which affect the biosorption are; agitation rate, pH, temperature, sorption time, sorbent dose, particle size and initial metal-ion concentration [1,2,5,6]. Among these process variables, the initial ion concentration, pH, temperature and particle size are found to be the predominant variables, these four variables are used as the input vectors to train the network whereas the other three variables are kept constant at their optimum values.

- Variables
  Initial Cu ion concentration (mg/l): 50, 60, 70 and 80
  pH: 3–6;
  temperature (°C): 25, 30, 35 and 40;
  particle size (μm): 50, 100, 150 and 200.
- Constants
  Agitator speed: 200 rpm;
  adsorption time: 90 min;
  sorbent dosage: 500 mg/50 ml;

3.2. Experiments

An experimental design for the biosorption of Cu(II) using sawdust has been developed by assigning four different experimental modes as listed, here each time three of the variables were kept constant and the remaining one is altered within the selected range:

- Mode 1: In this mode of operation, the initial ion concentration was varied keeping the pH, temperature and particle size constant.
- Mode 2: In this mode of operation keeping the initial ion concentration, temperature and particle size constant and varying the pH generated the data.
- Mode 3: In this mode of operation, temperature was varied keeping the other variables constant.
- Mode 4: In this mode of operation, the data are generated by keeping all the variables constant except the particle size.

As the process has four variables with four levels, using their factorial design the experimental data (256 experimental runs) were generated. For each experimental run the amount of copper biosorbed was estimated and thus the biosorption efficiency.

3.3. Batch adsorption studies

1. The aqueous solution of copper chloride having desired copper ion concentration (50, 60, 70, 80 mg/l) was taken and its pH is adjusted to desired value.
2. To the 50 ml of above solution 500 mg of particles having the size of 50, 100, 150, 200 μm is added in separate conical flasks.
3. These flasks are then placed in the orbital shakers under controlled temperatures (25, 30, 35 and 40 °C).
4. The speed of the orbital shaker is fixed at 200 rpm and adsorption is carried out until equilibrium is achieved, i.e., 90 min.
5. After attaining equilibrium, contents of the flasks are filtered separately and then analyzed for the amount of copper remaining in the solution.

Table 1
Network topology

Architecture	Recurrent (Elman)
Input processing elements	4
Output processing elements	1
Exemplars	4864
hidden layers	3
input layer	Axon
Recurrence	Partially recurrent
Hidden layer 1	
Processing elements	50
Transfer function	TanhAxon
Learning rule	Momentum
Step size	0.010
Momentum	0.700
Hidden layer 2	
Processing elements	40
Transfer function	TanhAxon
Learning rule	Momentum
Step size	0.10
Momentum	0.700
Hidden layer 3	
Processing elements	27
Transfer function	TanhAxon
Learning rule	Momentum
Step size	0.010
Momentum	0.700
Output layer	
Processing elements	1
Transfer function	Linear Sigmoid Axon
Learning rule	Momentum
Step size	0.0010
Momentum	0.700
Learning	Supervised learning
Maximum epochs	1000
Termination	Mean square error Threshold 0.001, minimum
Weight update	Batch
Probe configuration	MSE: 0.002139579316, %error: 1.123533114529, elapsed time 2:38:13

4. Network design

4.1. Construction of the network

NeuroSolutions® 5.0 package (evaluation) with Pentium III 866 MHz computer was used for construction, training, testing, simulation and validation of the network.

A recurrent network (Elman) with three hidden layers, containing 50, 40, 27 neurons, respectively, was found to be appropriate for the system considered in this work. The topology of the network is presented (Table 1), and the five-layered partially recurrent network architecture is shown in diagram (Fig. 3). As per the network topology the neural network employed has four input nodes corresponding to the four process variables namely, initial ion concentration, pH, temperature and particle size. The corresponding experimental percentage adsorption efficiency is fed for training purpose.

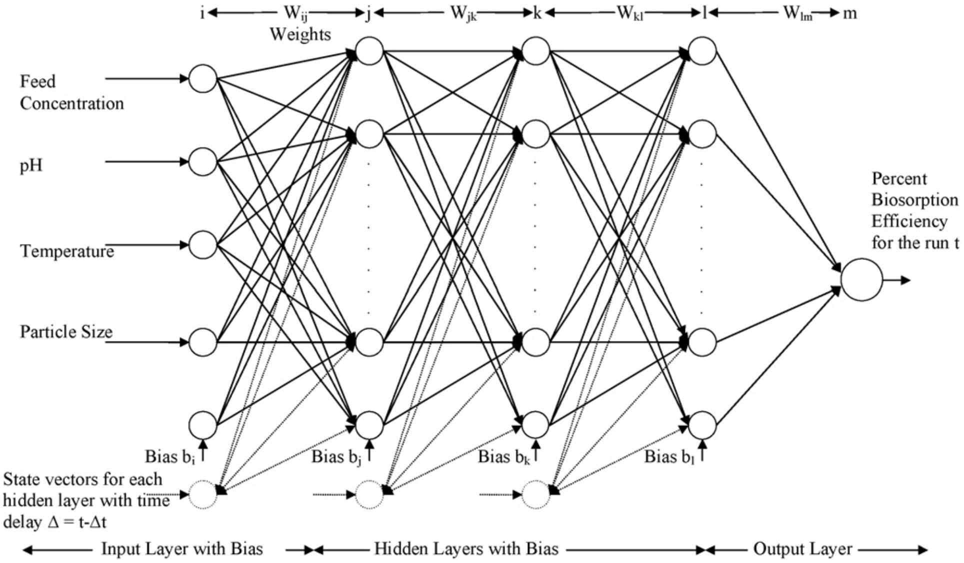

Fig. 3. Five-layered partially recurrent network architecture.

4.2. Data generation

As the experimental data obtained (256 numbers) are not sufficient to train the network, they were further smoothened using linear interpolation in order to produce sufficient data to train the network effectively within the selected range of each variable. This data generation is done by SPSS 10.0 (evaluation). Thus, the final lot produced consist 4864 data sets.

Two subsets of data are used to build the network model: a training set and a testing set. The training phase needs to produce a neural network that is both stable and convergent. In the testing phase the performance of the network is tested for unused experimental data obtained within the range of selected variables.

For simulation and validation of the network a separate 25 numbers of experiments were carried out; for each experiment the values of the variables are selected in such a way that they are within the selected range but randomly out of their levels, and the results are given in Table 2. This step ensures to validate the adoptability of the network for

Table 2
Simulation and validation data—biosorption of Cu(II) by sawdust

Experimental run	Feed concentration (mg/l)	pH	Temperature (°C)	Particle size (μm)	Experimental %efficiency	Simulated %efficiency	Relative error	Percentage relative error
1	55	3.5	27.5	75	88.2	87.431	−0.769	−0.879
2	55	3.5	27.5	160	84.3	84.574	0.274	0.324
3	59	4.6	39	80	81.1	81.541	0.441	0.541
4	64	4.3	34	140	83.4	84.360	0.960	1.138
5	68	5.3	32	175	78.5	78.525	0.025	0.032
6	73	5.8	29	60	87	85.904	−1.097	−1.299
7	78	3.8	38.5	130	76.3	76.430	0.130	0.170
8	71	4.9	35	195	85.5	84.436	−1.064	−1.260
9	63	3.2	26	88	86.1	85.202	−0.898	−1.053
10	67	4.5	32	125	82.2	82.751	0.551	0.665
11	61	5.3	38	160	79	79.605	0.605	0.760
12	53	3.3	28	150	85.2	85.260	0.060	0.070
13	59	3.8	31.4	190	80.4	80.797	0.397	0.485
14	64	4.7	37	65	71.5	72.408	0.908	1.255
15	58	4.5	25	135	86.4	85.849	−0.551	−0.642
16	69	5.2	34.3	110	82.2	81.561	−0.639	−0.783
17	74	5.6	36	50	83.4	82.533	−0.867	−1.051
18	57	4.4	31	115	81.1	81.258	0.158	0.190
19	64	3.9	33	135	77.5	78.788	1.288	1.635
20	78	3.3	40	190	74.6	75.209	0.609	0.800
21	52	5.1	28.5	95	87.2	87.606	0.406	0.463
22	67	5.9	30.5	105	83.2	82.194	−1.006	−1.224
23	77	4.2	33.7	185	79	78.982	−0.018	−0.023
24	58	3.2	26	85	83.4	84.336	0.936	1.110
25	65	5.8	34	160	85.3	85.160	−0.140	−0.160

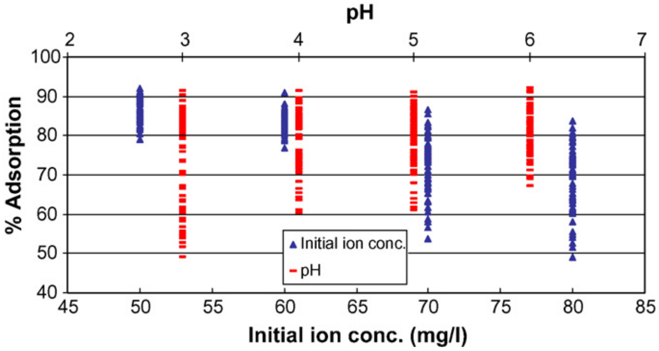

Fig. 4. Experimental trend for initial Cu^{2+} concentration and pH.

inputs out of the training data and their pre-defined levels.

5. Results and discussion

The experimental design has been made with the purpose of exploring the effect of interaction of the four variables namely feed concentration, pH, temperature and particle size on the adsorption efficiency, by implementing the permutations of variables within their selected range.

5.1. Observations from the experimental data

Even though the theory of the effect of variables on the adsorption efficiency is not required for neural network modeling, the observation may be helpful in understanding the complexity and non-linear behavior of the process, and thus the usefulness of ANN prediction.

The observed experimental trends are shown in Figs. 4 and 5.

5.1.1. Effect of initial ion concentration and pH

The scattered plot (Fig. 4) represents the trend in experimental %adsorption efficiency for the changes in initial Cu^{2+} ion concentration and pH. The clusters in the plot represent the four initial ion concentrations of Cu^{2+} (50, 60, 70 and 80 mg/l) and the four pH values (3–6). The trend indicates that the process is highly sensitive for any change in these two variables. The adsorption efficiency decreases against the increase in initial ion concentration. Clearly 50 and 60 mg/l concentrations approach the maximum removal, indicates lower initial concentrations help better adsorption.

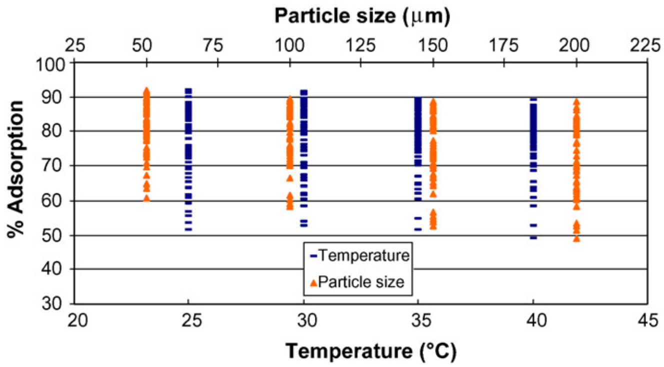

Fig. 5. Experimental trend for temperature and particle size.

The overall trend in pH shows high-degree of non-linearity, similar to the changes in initial ion concentration. As higher content of H^+ ions in the solution favors the adsorption process, better adsorption occurs for pH 6; probably, the effect of neutralized condition and adsorption under basic conditions need to be studied for further conclusions.

5.1.2. Effect of temperature and particle size

Fig. 5 represents the efficiency variations for temperatures (25, 30, 35 and 40 °C) and particle sizes (50, 100, 150 and 200 μm). Higher temperatures always favor the desorption process and here the trend reveals the same. In the case of particle size, as the overall surface area available for adsorption increases with decrease in particle size, here the adsorption efficiency decreases with increase in particle size. Even though low temperature and low particle size produce significant improvements in the adsorption process, comparatively they show less effect on the efficiency than the other two variables.

5.2. Performance of the network

Out of 4864 data sets, 4608 data sets were used to train the network and selected 256 data sets were used for testing of the

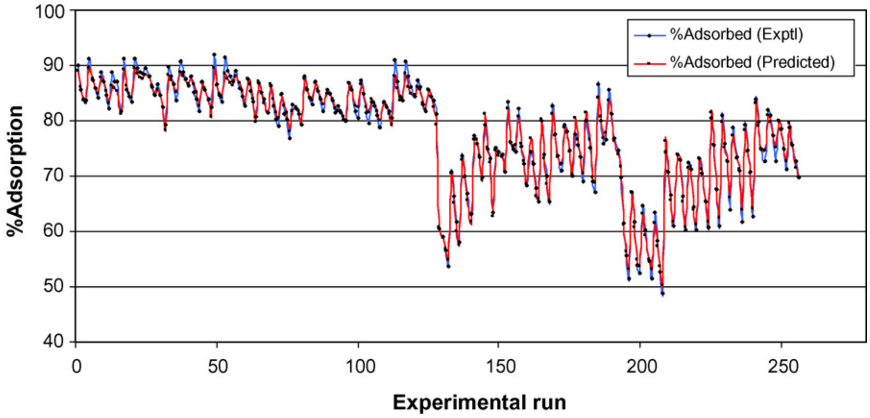

Fig. 6. Comparison between %adsorption (experimental) and %adsorption (predicted) using test data.

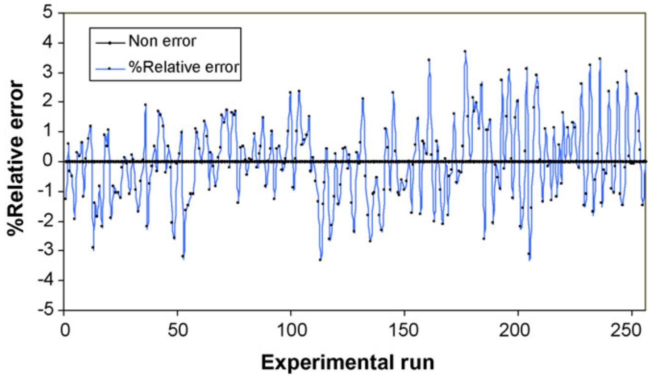

Fig. 7. Percentage relative error between test data and ANN predicted values vs. experimental run.

NN model. During training of the network, the least mean square error has been kept at 0.001 and the frequency of progress (in epochs) is set at 50 with maximum epochs of 1000 to train the network.

The performance of the network upon training with the test data may be well exhibited by the closely followed trends between the actual and predicted patterns in Fig. 6. The recurrent neural network is found to be very efficient in predicting the extent of adsorption within the range of data contained in the training set, i.e., 4608 data set.

Fig. 7 represents the percentage relative error between ANN testing (256 runs) and experimental data. The ultimate mean square error is found to be 0.002139579 which is within ±1% error range.

Fig. 8 represents comparison between ANN simulated (25 data sets) and experimental output data. The percentage relative error is between +1.635 and −1.299 (Table 2). The 25 data sets were purposely selected to remain random so as to validate the effectiveness of the network for untrained inputs. The introduction of recurrence over the back-propagation improves the forecasting performance of ANN in comparison to the use of back-propagation alone. The predicted adsorption efficiencies of ANN results obtained are as close as with training and testing data given to the network.

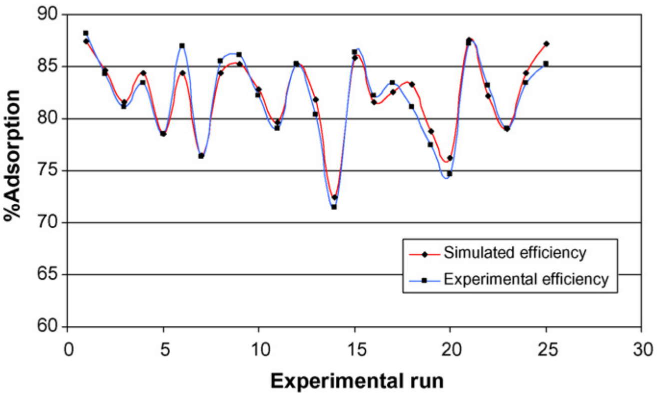

Fig. 8. Comparison between ANN simulated and experimental output data.

6. Conclusion

In the present investigation, a recurrent neural network has been designed and demonstrated to predict the extent of copper ion adsorption with sawdust from aqueous copper solutions by taking into account the effect of initial copper concentration, pH and temperature and particle size of the adsorbent. A simple back-propagation recurrent network using the momentum-training algorithm is found to be very effective to generalize and predict the degree of adsorption. The configuration of the recurrent neural network that gives the best prediction is the one with three hidden layers consisting of 50, 40 and 27 neurons in each layer. ANN predicted results are very close to the experimental values. The average mean square error is 0.002139579, which is sufficient to have error within ±1%. The present work suggests that neural network can be used as an effective technique in modeling, estimation and prediction of biosorption process. Also, neural network can be considered as an effective supplement for the conventional and complicated mathematical models in the prediction of bioprocess parameters.

Acknowledgement

The authors wish to express their gratitude for the support extended by the authorities of Annamalai University, Annamalai Nagar, India in carrying out the research work in Bioprocess Engineering laboratory, Department of Chemical Engineering.

References

[1] N. Chubar, J.R. Carvalho, M. Joana Neiva Correia, Cork biomass as biosorbent for Cu(II), Zn(II) and Ni(II), Colloids Surf. A: Physicochem. Eng. Aspects 230 (2003) 57–65.

[2] C.R.T. Tarley, M.A.Z. Arruda, Biosorption of heavy metals using rice milling by-products. Characterization and application for removal of metals from aqueous effluents, Chemosphere 54 (2004) 987 –995.

[3] M. Iqbal, R.G.J. Edyvean, Biosorption of lead, copper and zinc ions on loofa sponge immobilized biomass of *Phanerochaete chrysosporium*, Miner. Eng. 17 (2004) 217–223.

[4] G. Cetinkaya Donmez, Z. Aksu, A. Ozturk, T. Kutsal, A comparative study on heavy metal biosorption characteristics of some algae, Process Biochem. 34 (1999) 885–892.

[5] E. Susan Bailey, J. Trudy Olin, R. Mark Bricka, D. Dean Adrian, A review of potentially low-cost sorbents for heavy metals, Water Res. 33 (11) (1999) 2469–2479.

[6] M. Ajmal, A.H. Khan, S. Ahmad, A. Ahmad, Role of sawdust in the removal of copper(II) from industrial wastes, Water Res. 32 (1998) 3088 –3091.

[7] K. Pramanik, Use of ANN for predication of cell mass and ethanol concentration in batch fermentation using *saccbaromyces cerevisiae* yeast, Instit. Eng. (India) 85 (2004) 31–35.

[8] D.R. Baughman, Y.A. Liu, Neural Network in Bioprocessing and Chemical Engineering, Academic Press, San Diego, 1995.

[9] J.L. Elman, Finding structure in time, Cognitive Sci. 14 (1990) 179–211.

[10] X. Du, Q. Yuan, J. Zhao, Y. Li, Comparison of general rate model with a new model—artificial neural network model in describing chromatographic kinetics of solanesol adsorption in packed column by macro porous resins, J. Chromatogr. A 23 (2007) 165–174.

[11] J.S.J. van Deventer, S.P. Liebenberg, L. Lorenzen, C. Aldrich, Dynamic modelling of competitive elution of activated carbon in columns using neural networks, Miner. Eng. 8 (1995) 1489–1501.

[12] W. Gao, S. Engell, Estimation of general nonlinear adsorption isotherms from chromatograms, Comput. Chem. Eng. 29 (2005) 2242–2255.

[13] S. Haykin, Neural Networks: A Comprehensive Foundation, Prentice Hall, 1999.

[14] D. Mandic, J. Chambers, Recurrent Neural Networks, Wiley, 2001.

[15] http://en.wikipedia.org/wiki/Artificial_neural_network#Other_types_of_networks.