Prediction of Membrane Fouling Based on GA-RBF Neural Network and PCA

Yaoke Shi* and Zhiwen Wang

College of Electrical and Information Engineering, Lanzhou University of Technology, Lanzhou, China

*Corresponding author: yaoke_shi@163.com

Abstract. Aiming at the problem that membrane fouling cannot be accurately predicted due to the complexity of the influencing factors of membrane fouling, firstly, principal component analysis (PCA) is used to process membrane fouling data, and the three influencing factors that have the greatest impact on membrane flux are obtained. After processing the experimental data obtained, determine the structure of the radial basis function (RBF) neural network, and use the genetic algorithm (GA) to optimize the parameters of the RBF neural network, improve the network prediction accuracy, and train to establish a PCA-based GA-RBF membrane fouling prediction model proves through simulation that the optimized membrane fouling prediction model is superior to the traditional RBF neural network in terms of prediction accuracy and convergence speed, achieving the research purpose of accurately predicting the degree of membrane fouling

Keywords: Radial basis function, principal component analysis, genetic algorithm, membrane fouling.

1. Introduction

The wide application of membrane bioreactor (MBR) can effectively alleviate the shortage of water resources. However, membrane pollution in MBR process increases the feed pressure, reduces the productivity, increases the system downtime, and increases the operation cost, which seriously hinders the popularization and application of MBR process. The prediction of membrane fouling can timely replace and clean membrane modules, timely control feed, temperature, operating conditions and other factors, effectively inhibit membrane fouling, improve MBR production efficiency [1], reduce operating costs, and improve effluent quality. Due to the complexity of the influencing factors of membrane pollution, the prediction of membrane pollution is mostly based on the experience of skilled workers, which cannot complete the accurate prediction [2]. In recent years, the membrane pollution prediction model based on artificial neural network can complete the preliminary prediction of membrane pollution, and the prediction accuracy still cannot reach the ideal state [3]. In order to solve the problem that it is difficult to accurately predict membrane fouling due to the complexity of influencing factors of membrane fouling, principal component analysis (PCA) was used to process the data of membrane fouling, and the three influencing factors that have the greatest impact on membrane flux were obtained [4]. Then the experimental data were processed to determine the structure of RBF...
neural network and train RBF neural network, because the selection of network model parameters determines the prediction accuracy, genetic algorithm (GA) is used to optimize the parameters of RBF neural network, and the GA-RBF membrane pollution prediction model based on PCA is established [5]. The simulation results show that the optimized membrane pollution prediction model is superior to the simple RBF neural network in prediction accuracy and convergence speed, so as to improve the prediction accuracy of membrane pollution.

2. Radial basis function (RBF) neural network

RBF neural network is a kind of high performance feedforward neural network with three-layer structure and the best approximation property. The hidden layer of RBF neural network is composed of the "base" of hidden unit. The input is directly mapped to the hidden space without weight connection. The mapping relationship is determined by the corresponding RBF center point, and the relationship between hidden layer and input layer belongs to linear mapping.

The radial basis function neural network makes the signal enter the input layer to form the signal source node, then send it to the hidden layer for conversion, and then transmit it to the output layer in the form of input quantity. The output layer neuron outputs the product of the function value and weight of each layer [6]. The network structure is shown in Figure 1.

![Figure 1. RBF network structure.](image)

3. Genetic algorithm and its optimization

3.1. Realization steps of genetic algorithm

The stochastic calculation process of genetic algorithm is not without objective, but the optimal solution is obtained by calculating and operating chromosome adaptation value in the form of similar biological inheritance. The design of genetic algorithm usually includes the following aspects [7]:

- The coding scheme is determined. There are two coding methods, binary coding and real coding.
- Determination of fitness function. Fitness value is an important index of individual selection in the process of genetic algorithm, which is usually expressed as objective function and genetic probability.
- Select the control parameters. There are mainly population size, maximum algebra of evolution, probability of genetic operation and other auxiliary parameters.
- Genetic operators and selection strategies.

Selection operation: select by calculating fitness value, including local selection, tournament selection and proportion selection.

Crossover operation: select the same crossover point between two parent chromosomes, replace and recombine to get the offspring chromosome.

Mutation operation: there are gene mutations in the population, so that the algorithm can maintain diversity and local search ability.
• The termination condition of the algorithm. Set the maximum evolution algebra or the algebra of the algorithm when the fitness value of the solution does not change significantly, and the operation ends.

The flow of genetic algorithm is shown in Figure 2.

![Flow chart of genetic algorithm.](image)

**Figure 2.** Flow chart of genetic algorithm.

### 3.2 Optimization implementation steps

The RBF neural network is optimized by genetic algorithm.

- **Step 1**, when encoding the parameters of RBF neural network, the training speed and error are coded in real numbers, the maximum number of neurons and the interval between neurons are set as integer codes, and all the coded gene sequences are called an individual.

- **Step 2**, set randomly generated 20 individuals as the initial population to construct the initial mating pool.

- **Step 3**, train 20 individuals randomly in turn, calculate the prediction error, and specify the fitness of each individual as the calculated error percentage.

- **Step 4**, the roulette method is used to select the individual with the best survival ability for reproduction and eliminate other individuals.

- **Step 5**, according to the established crossover and mutation rate of genetic operation, 20 individual gene crossover and mutation, to obtain a new population, which can ensure the gene diversity.

- **Step 6**, repeat step 3, until the completion of twelve ethnic evolution, the end of the cycle, the individual is the best individual after optimization.
Step 7, after the training, the gene string on the best individual is the best parameter value for RBF network training. The four selected parameters can be used for MBR membrane pollution simulation test, and the best prediction effect can be obtained.

4. GA-RBF simulation model of membrane fouling in MBR

4.1. Acquisition and preprocessing of experimental data

The experimental data are from the membrane water treatment experiment of wet water treatment research center, Royal University of Melbourne, Australia. The experimental membrane module used in this paper is polyvinylidene fluoride (PVDF) hollow fiber microfiltration (MF) module. The water inflow mode is external pressure type, and the effective use area of the membrane is 20m$^2$. The membrane module is in the MBR reaction tank because of the microorganism, the water quality and the water quality the components of pollutants are complex, and the selection of influencing factors of membrane pollution largely determines the accuracy of membrane pollution prediction. Therefore, in order to improve the prediction accuracy of membrane pollution, the principal component analysis method is used to screen out the most influential factors of membrane pollution for experimental analysis. The introduction of principal component analysis can improve the prediction accuracy and reduce the prediction time. In order to ensure that the input and output data are in the same order of magnitude, the data is normalized first [8], then the input space is divided to facilitate the implementation of the algorithm. The MBR membrane pollution prediction model is built, which includes PCA and GA optimized RBF neural network. As shown in Figure 3.

![Figure 3. Simulation model design.](image)

4.2. Build GA-RBF model

1) Related parameter setting

The RBF neural network optimized by genetic algorithm is used to establish the membrane fouling prediction model of membrane bioreactor. The data are based on the experimental values of three factors which have been processed before. The parameters in this paper are determined by repeated experiments. In order to get the best experimental results, the initial population size is defined as $n = 20$, the maximum evolution algebra is set as $P = 12$, the mutation rate and crossover rate are 0.05 and 0.76 respectively, and the error range is $1e^{-4}$-$1e^{-3}$; The speed is 15-25; The maximum number of neurons was defined as 90-150; The interval between neurons was 10-30.

2) Training RBF neural network

Encoding the parameters to be optimized, 20 random individuals are generated, and each individual has four parameters. The difference between the training value of RBF neural network and the prediction sample is the fitness value of genetic operation. We set that the smaller the error percentage is, the smaller the fitness value is, and the better the individual is. Different from other genetic optimization training, the individual with the smallest error in the population will be recorded at the end of each training, and the cycle will continue from generation to generation.

3) RBF network optimized by GA

In the optimization process of genetic algorithm, the judgment of the qualified individuals and the optimal solution completely depends on the fitness function. The fitness function of GA in this paper is the percentage of experimental error. The code is as follows.
\[
\text{Delta} = \text{sum} \left( \text{abs} \left( \text{Output\_predict} - \text{Output\_out} \right) \right);
\]
\[
\text{Total} = \text{sum} \left( \text{abs} \left( \text{Output\_test} \right) \right);
\]
\[
\text{Error} = \frac{\text{Delta}}{\text{Total}};
\]

Random coding population generates genes, calculates fitness value and individual RBF prediction value, and then uses roulette method to select offspring with strong survival ability to continue inheritance. Roulette is executed for many times to select individuals. Each time, three random numbers between \([0, 1]\) are selected as pointers to select individuals suitable for genetic operation. Due to the randomness of roulette, individuals are occasionally repeatedly selected.

After 12 generations of population evolution, the error tends to be stable, and there is no big change. Set 12 as the maximum evolution algebra of genetic operation, and the optimal solution output is the individual with the minimum fitness value. Otherwise, continue to cycle until the above termination conditions are met.

4.3. Prediction results and comparative analysis

Divide 100 groups of data into 20 groups of prediction data and 80 groups of experimental data, set the maximum evolution algebra to 12, and compare the training results of optimized network and traditional RBF network, as shown in Figure 4.

![MBR membrane fouling prediction curve](image)

**Figure 4.** Comparison of prediction results.

The average error of RBF network training result is 0.2861, which can realize the simple prediction of membrane flux. Compare the relative error of the two kinds of network prediction, as shown in Table 1. Figure 5 shows the error curves of the two prediction networks. The average error of GA-RBF network is only 0.0256, and the prediction accuracy is significantly improved.
### Table 1. Prediction error analysis.

| sample | expectations | RBF predictive value | RBF relative error | GA-RBF predictive value | GA-RBF relative error |
|--------|--------------|----------------------|--------------------|-------------------------|------------------------|
| 1      | 25.6000      | 25.7674              | -0.1674            | 25.6104                 | -0.0104                |
| 2      | 24.6000      | 24.7841              | -0.1841            | 24.5927                 | 0.0073                 |
| 3      | 23.8000      | 23.6827              | 0.1173             | 23.8045                 | -0.0045                |
| 4      | 19.2000      | 18.6800              | 0.5200             | 19.2214                 | -0.0214                |
| 5      | 17.5000      | 16.5689              | 0.9311             | 17.5541                 | -0.0541                |
| 6      | 17.1000      | 17.2775              | -0.1775            | 17.0971                 | 0.0029                 |
| 7      | 16.5000      | 16.3962              | 0.1038             | 16.5212                 | -0.0212                |
| 8      | 19.6000      | 19.4296              | 0.1704             | 19.6102                 | -0.0102                |
| 9      | 22.2000      | 21.5841              | 0.6159             | 22.1894                 | 0.0106                 |
| 10     | 25.9000      | 25.8525              | 0.0475             | 25.9059                 | -0.0059                |
| 11     | 23.5000      | 23.6247              | -0.1247            | 23.4895                 | 0.0105                 |
| 12     | 18.8000      | 18.9544              | -0.1544            | 18.8143                 | -0.0143                |
| 13     | 17.5000      | 17.5578              | -0.0578            | 17.4756                 | 0.0244                 |
| 14     | 24.8000      | 24.9145              | -0.1145            | 24.8042                 | -0.0042                |
| 15     | 23.5000      | 24.5512              | -1.0512            | 23.4946                 | 0.0054                 |
| 16     | 20.4000      | 20.9947              | -0.5947            | 20.4049                 | -0.0049                |
| 17     | 21.9000      | 21.8154              | 0.0846             | 21.9748                 | -0.0748                |
| 18     | 20.9000      | 20.7681              | 0.1319             | 20.8865                 | 0.0135                 |
| 19     | 25.6000      | 25.8114              | -0.2114            | 25.5945                 | 0.0055                 |
| 20     | 18.5000      | 18.3373              | 0.1627             | 18.5137                 | -0.0137                |

![Graph](image.png)

**Figure 5. Prediction error curve.**

The error of GA-RBF prediction model tends to a very small stable value with the progress of genetic optimization, as shown in Figure 6. According to the previous setting, the maximum evolution algebra is 12. In the network training process, when approaching 12 generations, the prediction result is in a relatively stable range, which can jump out of the cycle and end the optimization process. The
GA-RBF neural network membrane pollution prediction model is simulated and verified. The experimental results show that the error of the prediction model is significantly reduced compared with the traditional RBF prediction model, and the performance is stable. The training results are significantly better than the results before optimization, and the prediction accuracy is ideal.

5. Summary
The intelligent characteristics of RBF neural network can predict MBR membrane pollution according to the characteristics of membrane flux. Because the prediction accuracy depends on the selection of network model parameters to a large extent, this chapter analyzes the main factors of membrane pollution and obtains the three main factors. The membrane pollution is characterized by membrane flux, the prediction model of RBF membrane pollution is constructed, and the relevant parameters of RBF network are optimized by genetic algorithm. The results show that the prediction accuracy of membrane pollution prediction model based on GA RBF is significantly improved compared with the single RBF model, and the experiment has achieved the expected goal.

Acknowledgments
This research was funded by the Natural Science Foundation of China under grant no. 61863026. Also, it is partially supported by the Industrial Support and Guidance Project for Higher Education of Gansu Province (2019C-05) and the Open Fund Project of Key Laboratory of Gansu Advanced Control for Industrial Process (2019KFJJ03).

References
[1] M. Lares, M.C. Neibi, and M. Sillanpaa, “Occurrence, identification and removal of microplastic particles and fibers in conventional activated sludge process and advanced MBR technology,” Water Res, vol. 133, pp. 236-246, 2018. (references)
[2] X.J. Du, Y.K. Shi, V. Jegatheesan, I.U. Haq, “A review on the mechanism, impacts and control methods of membrane fouling in MBR system,” Membranes, vol. 10, 2020. (references)
[3] V. Jegatheesan, B.K. Pramanik, J.Y. Chen, D. Navaratna, C.Y. Chang, L. Shu, “Treatment of textile wastewater with membrane bioreactor: A critical review,” Bioresour. Technol. vol. 204, pp. 202–212, 2016. (references)
[4] S. Chellam, “Artificial neural network model for transient crossflow microfiltration of polydisperse suspensions,” J. Membr. Sci. vol. 258, pp. 35–42, 2005. (references)
[5] H.D. Park, Y.H. Lee, H.B. Kim, “Reduction of membrane fouling by simultaneous upward and downward air sparging in a pilot-scale submerged membrane bioreactor treating municipal wastewater,” Desalination, vol. 251, pp. 75–82, 2010. (references)
[6] X.X. Yang, C.Q. Li, “Research of MBR simulation predictions based on improved extreme
learning machine. Comput,” Eng. Softw. Vol. 37, pp. 17–20, 2016. (references)

[7] Z. Wang, Y.F. Wang, “Creep rupture life estimation of P91 steel pipes through artificial neural network based hardness prediction,” J. Chin. Soc. Power Eng. Vol. 40, pp. 936–939, 2020. (references)

[8] O. Sudakov, E. Burnaev, D. Koroteev, “Driving digital rock towards machine learning: Predicting permeability with gradient boosting and deep neural networks,” Comput. Geosci., vol. 127, pp. 91–98, 2019. (references)