Estimating the Relative Tray Efficiency of Sieve Distillation Trays by Applying Artificial Neural Networks

X.Y. You* and Z. S. Yang **
*School of Environmental Science and Engineering,
Tianjin University, 300072 Tianjin, P.R. China.
**School of Mechanical Engineering,
Tianjin University, 300072 Tianjin, P.R. China

Original scientific paper
Received: February 28, 2002
Accepted: March 15, 2003

Artificial Neural Network (ANN) is applied to estimate the relative tray efficiency of sieve distillation trays. The training database is obtained from the results of *You* et al. (2001).[1] The feed-forward artificial neural network is adopted and trained by the back-propagation algorithm. 150 sets of data is used to train and test the network. The results show that ANN model with one hidden layer gives a very close estimation of the relative tray efficiency.

Keywords
Relative tray efficiency, artificial neural network (ANN), back-propagation (BP) algorithm

Dedicated to: Professor Dr-Ing K.T. Yu for the celebration of his 80th birthday.

Introduction

Artificial Neural Network (ANN) has been widely applied in many science and engineering fields in recent years. It is a highly simplified and approximated network for Biological Neural Network (BNN) and furthermore, it can resemble the working pattern of human brain. The main advantages of ANN are economy, fast computing, easy realization, and independence from physical model compared with the traditional methods, such as experiment and numerical simulation. McCulloch and Pitts first introduced M-P model, which is considered as the origin of ANN, in 1943. After that, ANN underwent a relatively slow-developing period due to the lack of theory at first and efficient numerical algorithm later. Since the introduction of back propagation (BP) algorithm by *Rumelhart* et al.[2] in 1986, ANN has developed rapidly and its application has blossomed in many fields of science and industry.

Distillation is a widely used method to separate liquid mixtures into their components and has been applied to the separation processes in petroleum, petrochemical, chemical and related industries etc. It is commonly acknowledged that distillation is a very important process in today's industry. With the rapid development of computational fluid dynamics (CFD), the selection of tray parameters is usually depending on the theoretical analysis instead of experience or experiment. Several theoretical models were presented during last few years. *Liu* et al.[3] and *Yu* et al.[4] formulated two-dimensional models with single-fluid (liquid) theory and two-fluid (liquid and vapor) theory to describe the flow patterns on tray, respectively. *You* et al.[1] simulated the two-phase flow on trays by adopting a two-dimensional mixed two-phase flow theory and, furthermore, they showed that the relative tray efficiency can be increased as a result of improving the flow pattern on a tray. *Krishna* et al.[5] and *Baten* and *Krishna*[6] presented a three-dimensional CFD model, and computed by using CFX software for simulating flow field on sieve trays. Even though the above studies show lights on designing high performance trays, the requirement of large computer resources precludes CFD method from applying to a real distillation design, especially for the case of three-dimensional CFD simulation. With the concept of ANN, it is natural to use the operational and geometrical parameters of the tray as the input variables of ANN to predict the relative tray efficiency. For applying ANN, a database that includes all necessary information should be set up. The database can be obtained by either experimental study or CFD simulation. In this paper, the database is set up by the CFD study of *You* et al.[1]. Therefore, the aim of this paper is to develop an ANN model that can be applied to predict the relative tray efficiency.

Some studies on the application of ANN to distillation process appeared in recent years. *MacMurray* and *Himmelblau*[7] used ANN to model and control a packed distillation column, where the column ex-

*Corresponding author, email:xyyou@tju.edu.cn,
tel: + 86-22-27401141, fax: + 86-22-27401647.

hibits a change in the sign of the gain under various operating conditions. They showed the performance of the column could be modeled well by applying ANN. *Zamprogna* et al.[8] developed a virtual sensor to estimate the composition in a middle-vessel batch distillation column. The sensor was based on a recurrent artificial neural network and used the information available from secondary measurements (such as temperature and flow rate). The results indicated that the estimated compositions are in good agreement with the actual values.

In this approach, an artificial neural network is designed to predict the relative tray efficiency. The network is trained by the database based on the CFD results of *You* et al.[1] For the database based on the other results, ANN model can be trained by the same way.

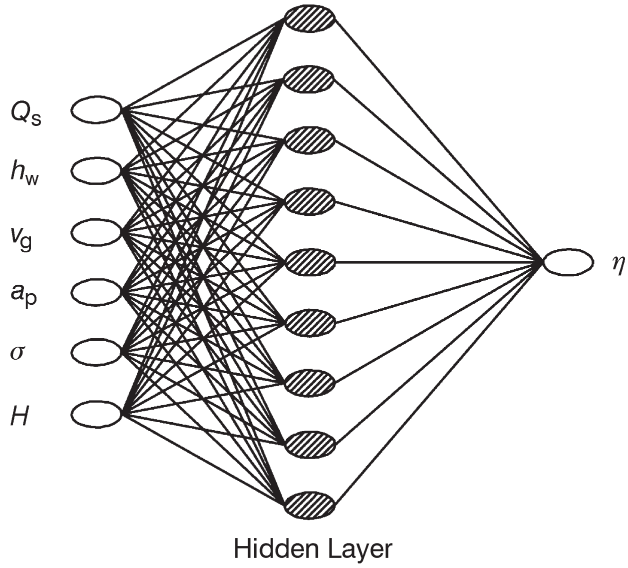

Fig. 1 – *A Simple 6-9-1 neural network, the lines connecting the neurons represent the weights*

Database generation

The results of *You* et al.[1] are used to establish the database. 150 sets of data are adopted to train and test the ANN model. In this study, the six quantities, such as liquid flow rate Q_s, outlet weir height h_w, superficial gas velocity $\boldsymbol{v}_g$, free area rate of sieve tray α_p, surface tension σ, Raoult's law constant H, have effects on the relative tray efficiency and they are used as the input parameters to predict the relative tray efficiency (the only output parameter).

ANN design and training

A feed-forward artificial neural network trained by back-propagation algorithm, is widely used. Based on the complexity of the problem and the size of the database, the number of hidden layers and the neurons within each hidden layer, can be varied. Figure 1 shows a 3-layer ANN structure with six inputs, one hidden layer with 9 neurons, and one output. The input layer that distributes the inputs to the hidden layer does not have any activation function.

Mathematically the network computes; see the ANN structure shown in Figure 1,

1. The output of the hidden layer (treating the bias as another input, but being not counted as a real input of our ANN structure)

$$h1(j) = \mathrm{Sum}(w1(i,j)^*\mathrm{in}(i),\ i = 1,7)$$

$$s(j) = f(h1(j))$$

2. For the output layer calculation

$$h2(k) = \mathrm{Sum}(w2(j,k)^*s(j),\ j = 1,9)$$

$$O(k) = f(h2(k))$$

Where, in(i) is the network input, $O(k)$ is the network output, $w1(i,j)$ represents the weight connecting neutron i in the input layer to neutron j in the hidden layer, $w2(i,j)$ represents the weight connecting neutron j in the hidden layer to neutron k in the output layer, and $f(x)$ is the neutron transfer function, for example a sigmoid $f(x) = 1/(1+\exp(-x))$, which is adopted in this study.

Training an ANN involves using a database of examples, which has values for the input and output of ANN. The ANN would learn by adjusting the weights to minimize the error of the outputs.

Selection of the number of hidden layer

The hidden layer abstracts the characters of the input information. Increasing the number of hidden layers, the performance of the treatment ability of a feed-forward artificial neural network will be promoted. But it may cause the complexity of training procedure, the increase of training samples and training time. Generally speaking, we should start with a system with one hidden layer and increase the number of the hidden layer by requirement. For a continuous output or even a discontinuous output in some cases, a very close prediction can be obtained via a feed-forward artificial neural network with only two hidden layers. In most cases, even a feed-forward artificial neural network with one hidden layer, can reach a considerable good prediction.

Selection of the number of hidden layer neuron

The selection of the number of hidden layer neuron is very important and troublesome. If the number of the hidden layer neuron is fewer, the ANN cannot receive all necessary information of the modeling system and has less tolerance on faults, so that it gives wrong outputs. On the contrary, the ANN may cause a phenomenon called overfitting. The overfitting is that the ANN can even recognize the noise during the training procedure. Under this circumstance, the outputs of ANN are ideal when the training samples are applied, while the outputs are not satisfied when the test samples are adopted. At the same time, ANN needs a long time for training and running. In some cases, it may even make it hard to determine the weights. After studying the ANN model with one hidden layer, Hecht-Nielsen pointed out that the number of hidden layer neuron should be $2N + 1$ to obtain reasonable output for any inputs. Here N is the number of input neuron. Even if there are some theories to determine the optimal number of hidden layer neuron, see *Kung* and *Hwang*[9] and *Kung* and *Hu*[10] for details, the selection of the number of hidden neuron depends mostly on experience. In this paper, the networks with different hidden layer neuron are simulated. The optimal network is found according to the RMS error and the ratio of eligible described in next section.

Procedure of training ANN

After defining the structure of ANN model, data are then collected and fed to the model. The network is trained to recognize the relationship between the input and the output parameters. The input layer distributes the inputs to the hidden layer. The lines connecting the neurons represent the weights. At the beginning of trainings, the weights of the network are randomly chosen. For a fast convergence, the initial weights are in the range of (–1,1). As the training procedure starts, an algorithm to minimize the difference between the network-predicted and the desired output adjusts the network parameters, such as the weights. The back propagation algorithm is used for this purpose since it is one of the most popular and extensively used algorithms for network training. Back propagation is a kind of a steepest descent method of optimization. Although it has been successfully applied in many fields, it suffers some shortcomings, such as slow convergence in some cases. Some authors prefer other optimization algorithms rather than back propagation, for example Levenberg-Marquardt method[11] and Quasi-Newton learning method.[12] However, the back propagation works well for our case and it is chosen to train our networks.

By using back propagation, the network learns through an iterative procedure, involving two steps performed many times. First, the examples of training data shown to the network are passed forward to the output layer to compute the errors at the output. The second step works backward through the network. The errors at the output layer are propagated backwards through the network and the weights allocated to each neuron connection are adjusted to minimize the error in the output data. Using this technique it is possible for the network to become trapped in a local minima. For this reason, a supervised training method was used. After the training data have been presented to the network for a pre-determined number of times, a test data set are presented. The results from the previous presentation of the test dataset is compared, and, if there is an improvement, the training continues. The cycle of presenting training and testing data continues until no improvement has been noted with the test dataset for 30 consecutive attempts. At this point the training is terminated. This prevents the network over-learning the training data by checking the performance with the test data and reduces the likelihood of the network to a local minimum as opposed to the global minimum by continuing training for 30 cycles after a minimum error has been achieved.

Once the network has been trained, a further validation set of data, which have not been used during training, is presented to ANN and the output compared with the known output in order to assess the predictive capabilities of the network.

Evaluation of ANN performance

The objective function to evaluate the performance of network is described as:

$$E = \sqrt{(\sum_{p=1}^{P} E_{\mathrm{p}})/p} \tag{1}$$

where, $E_p = \sum_{j=1}^{N} (t_{pj} - O_{pj})^2$, t_{pj} is the desired output of neuron j for pattern p. O_{pj} is the ANN output of neuron j for pattern p. If the target RMS error is lower than an anticipated value, ANN training is terminated.

Determination of ANN learning parameter

Automatically increase or decrease the learning parameter according to the training progress. If the RMS error is decreasing, the learning parameter is increased, however if the RMS error is increasing, the learning parameter is decreased. In our program, the learning parameter is increased or decreased by multiplying its value with a parameter *cd*. As a general rule, the parameter *cd* should be a few percent different from 1 and the decreasing percent of the learning parameter should be larger than the increasing percent of the learning parameter. Based on this rule, the default values of parameter *cd* could be 1.02 and 0.96, which gives an increasing of 2 % and decreasing of 4 % for the learning parameter, correspondingly. For Liu tray, the learning parameter is 0.3153 for our well-trained ANN.

Our ANN models

In the model of *You* et al.[1], six quantities (i.e. liquid flow rate Q_s, outlet weir height h_w, superficial gas velocity $\boldsymbol{v}_g$, free area rate of sieve tray α_p, surface tension σ, Raoult's law constant H) are major factors to determine the relative tray efficiency. Thus, we choose the input neuron number as six. The anticipated output is only the relative tray efficiency. Then the output has one neuron.

Feed-forward artificial neural networks trained by BP algorithm, are used in this study. Based on the conclusion of Section 3, a feed-forward ANN model with two hidden layers can well predict continuous outputs and even discontinuous outputs corresponding to their input parameters. Thus, two kinds of ANN models are chosen to train and test here. One model consists of an input layer of six neurons corresponding to the six input parameters, one hidden layer and an output layer of one neuron representing the output parameter. Another model has the same structure except having two hidden layers. By comparing the results of ANN models, which have different number of neurons in each hidden layer, the optimal ANN structure can be obtained.

Results and discussions

ANN models, with either one hidden layer or two hidden layers, are trained and tested to look for an optimal ANN structure. In all studies, 120 and 30 sets of data are used to train and test ANN models, respectively. Each set of data had 7 parameters.

Determination of the optimal ANN structure

The following shows how to find the optimal ANN on Liu Tray (the small diameter tray). The geometric quantities for Liu tray are d = 1.2 m and b/d = 0.645. d and b are diameter and the width of the outlet weir, respectively. Table 1 shows the results of training and testing ANN model with one hidden layer.

In Table 1, 6-i-1 represents the ANN model with 6 input neurons, i neurons in the hidden layer and 1 output neuron. The ratio of eligible represents the rate of the number of samples with relative target error less than a presumed number (here 1 %) with respect to that of all samples. It represents the ratio of ANN output, which satisfies the anticipation. The results of Table 1 indicate the following points:

Table 1 – *The results of training and testing ANN model with one hidden layer*

	Average training results			Average testing results	
ANN Structure	Iterations	RMS Error	Ratio of Eligible	RMS Error	Ratio of Eligible
6-2-1	3515	0.010878	75 %	0.007783	76.7 %
6-3-1	3590	0.005940	90 %	0.003671	93.3 %
6-4-1	5425	0.006089	90 %	0.003826	86.7 %
6-5-1	3145	0.004700	95 %	0.003386	100 %
6-6-1	5055	0.005630	88 %	0.003345	93.3 %
6-7-1	6160	0.003039	98 %	0.001782	100 %
6-8-1	4040	0.003692	98 %	0.002562	100 %
6-9-1*	2870	0.003231	98 %	0.002025	100 %
6-10-1	2545	0.003990	98 %	0.002083	96.7 %
6-11-1	3590	0.003961	98 %	0.002649	96.7 %

– Both, the RMS errors of training and testing, decrease when the number of hidden layer neurons changes from 2 to 9 or from 11 to 9.

– The ratio of eligible becomes nearly 100 % for the structures whose number of hidden layer neuron is larger than 6.

– The least iteration occurs at the ANN structure 6-10-1. But its ratio of eligible of testing is less than that of structure 6-9-1.

RMS Error and the ratio of eligible are two major factors to evaluate the performance of ANN model. RMS should be small enough and the ratio of eligible should be high enough. From Table 1, it is concluded that ANN structure 6-9-1 is the optimal one. Figure 1 shows the ANN structure 6-9-1.

Now we turn to study the performance of ANN with two hidden layers. The results are presented in Table 2.

In Table 2, 6-i-j-1 represents the ANN model with 6 input neurons, i and j hidden layer neurons respective for the first and second hidden layer and 1 output neuron. Table 2 indicates:

– Although, some networks with two hidden layers converge fast comparing with the network with one hidden layer, they cannot have the high ratio of eligible.

– When the number of neuron in the first hidden layer is more than 8, the connecting weights are more than 77. Thus ANN needs a much longer time

Table 2 – *The results of training and testing ANN model with two hidden layers*

	Average training results			Average testing results	
ANN Network	RMS Error	Iterations	Ratio of Eligible	RMS Error	Ratio of Eligible
6-3-2-1	0.006355	3450	87 %	0.003959	91.7 %
6-4-2-1	0.005765	3500	90 %	0.003825	100 %
6-4-3-1	0.006153	3100	87 %	0.003955	100 %
6-5-2-1	0.004175	4186	95 %	0.002584	100 %
6-5-3-1	0.004354	5790	97 %	0.002499	96.7 %
6-6-2-1	0.003764	3330	97.5 %	0.002129	100 %
6-7-2-1	0.005918	1715	91 %	0.003999	93.3 %

for training. It is not economic and will not be adopted as optimal ANN structure here.

– Two hidden layer ANN network, which spends comparable training time as the optimal one hidden layer ANN 6-9-1, shows no improvements comparing to that with one hidden layer.

After comparing Tables 1 and 2, ANN network with structure of 6-9-1 is selected to simulate the relative tray efficiency.

Further testing for optimal ANN

It is shown above that the well-chosen ANN network can model the relative tray efficiency of Liu tray. 12 typical samples are used to test further the trained optimal network. The results are given in Table 3. The RMS Error is required to be less 0.20 % and the ratio of eligible is 100 %, for the target error is less 1 %.

The optimal ANN structure 6-9-1 for Liu tray is also used for Porter tray. The geometric quantities for Porter tray are diameter d = 2.44 m and b/d = 0.615. The Porter tray has a larger diameter and a smaller b/d comparing to those of Liu tray. The results of training are the iteration 3065, the RMS error 0.007948, and the ratio of eligible 83 % for the target error 0.01. It is found that the ratio of eligible for Porter tray is only 83 %, which is much lower than that for Liu tray. It is due to the different diameters of trays. It is clearly shown in *You* et al.[1] that the tray diameter has strong effects on the flow pattern on the tray and, further, on the relative tray efficiency. This suggests the diameter should be an input parameter. This is a topic of further research.

30 samples are further used to test the trained network of Porter tray. The results of 12 typical cases are given in Table 4. Here, the RMS Error is required to be less than 0.20 %. The ratio of eligible is 90 % for the target error is 1 %. It is found that the largest target error is only 1.74 %. Thus, it is quite reasonable to conclude that the results of optimal ANN structure 6-9-1 for Porter tray are, also acceptable. This conclusion also indicates that the optimal ANN structure may not be sensitive to the diameter of tray.

Conclusions

The aim of present research is to determine, whether, ANN model can accurately predict the rel-

Table 3 – *Test results of the optimal ANN structure 6-9-1 on Liu tray*

No	Q_s	h_w	$\boldsymbol{v}_g$	α_p	σ	H	ANN Output	Efficiency[1]	Error
01	0.012	0.02	1.5	0.1	0.1	2.0	1.276	1.287	– 0.85 %
02	0.024	0.02	1.5	0.1	0.1	2.0	1.102	1.106	– 0.36 %
03	0.02	0.014	1.5	0.1	0.1	2.0	1.131	1.138	– 0.62 %
04	0.02	0.031	1.5	0.1	0.1	2.0	1.178	1.177	+ 0.08 %
05	0.02	0.02	0.75	0.1	0.1	2.0	1.057	1.060	– 0.28 %
06	0.02	0.02	1.9	0.1	0.1	2.0	1.153	1.152	+ 0.09 %
07	0.02	0.02	1.5	0.09	0.1	2.0	1.139	1.131	+ 0.71 %
08	0.02	0.02	1.5	0.19	0.1	2.0	1.132	1.131	+ 0.09 %
09	0.02	0.02	1.5	0.1	0.018	2.0	1.641	1.630	+ 0.67 %
10	0.02	0.02	1.5	0.1	0.11	2.0	1.128	1.120	+ 0.71 %
11	0.02	0.02	1.5	0.1	0.1	0.73	1.120	1.119	+ 0.09 %
12	0.02	0.02	1.5	0.1	0.1	1.9	1.132	1.131	+ 0.09 %

Table 4 – *Test results of the optimal ANN structure 6-9-1 on Porter tray*

No	Q_s	h_w	$\boldsymbol{v}_g$	α_p	σ	H	ANN Output	Efficiency[1]	Error
01	0.012	0.02	1.5	0.1	0.1	2.0	2.226	2.217	+ 0.36 %
02	0.024	0.02	1.5	0.1	0.1	2.0	1.369	1.383	– 1.01 %
03	0.02	0.014	1.5	0.1	0.1	2.0	1.412	1.411	+ 0.07 %
04	0.02	0.033	1.5	0.1	0.1	2.0	1.666	1.658	+ 0.48 %
05	0.02	0.02	0.75	0.1	0.1	2.0	1.128	1.148	– 1.74 %
06	0.02	0.02	1.9	0.1	0.1	2.0	1.683	1.672	+ 0.66 %
07	0.02	0.02	1.5	0.084	0.1	2.0	1.576	1.552	+ 1.55 %
08	0.02	0.02	1.5	0.18	0.1	2.0	1.350	1.362	– 0.88 %
09	0.02	0.02	1.5	0.1	0.017	2.0	3.891	3.845	+ 1.20 %
10	0.02	0.02	1.5	0.1	0.17	2.0	1.286	1.276	+ 0.78 %
11	0.02	0.02	1.5	0.1	0.1	0.72	1.375	1.391	– 1.15 %
12	0.02	0.02	1.5	0.1	0.1	1.6	1.448	1.454	– 0.41 %

ative tray efficiency for a wide range of geometrical, physical and operating parameters. It has been demonstrated that the optimal model is a network with one hidden layer. The application of ANN to the relative tray efficiency prediction indicates the coming of a flexible tool for engineers. Further work (experiment or CFD simulation) is required to provide a more completed database to train the network and validate its usefulness.

ACKNOWLEDGEMENT

The authors would like to acknowledge the research support by the Distillation Laboratory of the State Key Associated Laboratories of Chemical Engineering (China) at Tianjin University. One of the authors, X. Y. You, would also like to thank the invitation to the cooperative research.

Notation

d – tray diameter, m

H – constant put in Raoult's law

h_w – outlet weir height, m

Q_s – liquid flow rate per unit width, $m^2\ s^{-1}$

v_g – superficial gas velocity, $m\ s^{-1}$

b – width of outlet weir, m

Greek symbols

α_p – free area rate of sieve tray

η – relative tray efficiency

σ – surface tension, $N\ m^{-1}$

References

1. *You, X. Y., Yu, K. T., Yuan, X. G.*, The simulation of turbulent two-phase flow and mass transfer efficiency of sieve distillation trays, submitting for publication, 2002.
2. *Rumelhart, D. E., Hinton, G. E., Williams, R. J.,* Learning internal representations by error propagation in parallel distributed processing, Vol.1, Eds. Rumelhart, D. E. and McClelland, J. L., Cambridge, MIT press, 1986.
3. *Liu, C. J., Yuan, X. G., Yu, K. T.,* Chem. Eng. Sci., **55** (2000) 2287.
4. *Yu, K. T., Yuan, X. G., You, X. Y., Liu, C. J.,* Chem. Eng. Res. Des., **77** (1999) 554.
5. *Krishna, R., van Baten, J. M., Ellenberger, J., A. P. Higler, R. Taylor,* Chem. Eng. Res. Des., **77** (1999) 639.
6. *van Baten, J. M., Krishna, R.,* Chem. Eng. J., **77** (2000) 143.
7. *MacMurry, J. C., Himmelblau, D. M.,* Computers Chem. Eng., **19** (1995) 1077.
8. *Zamprogna, E., Barolo, M., Seborg, D. E.,* Trans. IchemE, part A, **79** (2001) 689.
9. *Kung, S. Y., Hwang, J. N.,* IEEE, (1991) I:363.
10. *Kung, S. Y., Hu, Yu Hen,* IEEE, (1991) II:163.
11. *Bulsari, A. B., Palosaari, S.,* Neural Computers and Applications, **1** (1993) 160.
12. *Chouai, A., Cabassud, M., Le Lann, M. V., Gourdon, C., Casamatta, G.,* Chem. Eng. Processing, **39** (2000) 171.