PROCESS MODEL
Chinese Journal of Chemical Engineering, **20**(6) 1142—1147 (2012)

A New Selective Neural Network Ensemble Method Based on Error Vectorization and Its Application in High-density Polyethylene (HDPE) Cascade Reaction Process*

ZHU Qunxiong (朱群雄)**, ZHAO Naiwei** (赵乃伟) **and XU Yuan** (徐圆)**
College of Information Science & Technology, Beijing University of Chemical Technology, Beijing 100029, China

Abstract Chemical processes are complex, for which traditional neural network models usually can not lead to satisfactory accuracy. Selective neural network ensemble is an effective way to enhance the generalization accuracy of networks, but there are some problems, *e.g*., lacking of unified definition of diversity among component neural networks and difficult to improve the accuracy by selecting if the diversities of available networks are small. In this study, the output errors of networks are vectorized, the diversity of networks is defined based on the error vectors, and the size of ensemble is analyzed. Then an error vectorization based selective neural network ensemble (EVSNE) is proposed, in which the error vector of each network can offset that of the other networks by training the component networks orderly. Thus the component networks have large diversity. Experiments and comparisons over standard data sets and actual chemical process data set for production of high-density polyethylene demonstrate that EVSNE performs better in generalization ability.
Keywords high-density polyethylene modeling, selective neural network ensemble, diversity definition, error vectorization

1 INTRODUCTION

Actual chemical processes usually present characteristics of nonlinearity and uncertainty, which make the modeling very complex, so that traditional neural network models can not lead to satisfactory accuracy. Neural network ensemble method is usually used in these processes to improve the generalization accuracy of networks. Hansen and Salamon initiate the neural network ensemble (NNE) in 1990, which trains a finite number of networks and combines the results together [1, 2]. The generalization accuracy of NNE depends on the accuracy of component networks and the diversity among them [3]. Therefore, how to enlarge the diversity among component networks is a main field in NNE research [4, 5]. Bagging [6] is a prevailing NNE training method based on bootstrap sampling [7], which trains the component networks with different sub-sets. Zhou *et al* [8, 9] proposed selective ensemble, i.e., selecting several of the component networks to make the ensemble (SEL-NNE) performs better than combining all of them (ALL-NNE). In general, a selective neural network includes training a number of networks and selecting some of them to combine the ensemble. Many researchers started to work on selective ensemble [10−13].

Since there is no unified definition of diversity among component networks, it is not convenient to select. Also, it is difficult to improve the accuracy of ensemble through selection when the diversities of available networks are all small, even if the selection method is good enough. Thus training the component networks with large diversity is the promise. However, the component networks trained by traditional method such as Bagging are independent, and the component networks may remain large correlation, which will lead to small diversity.

Aiming at the above problems, an error vectorization based selective neural network ensemble (EVSNE) method is proposed in this study. This method can train component networks with large diversities, since the output error vector of each component network can offset that of the other networks. Standard data sets are used to prove the improvement on the generalization accuracy of neural networks (NNs). Finally, the method is used in HDPE (high-density polyethylene) process.

2 ERROR VECTORIZATION

2.1 Definition of diversity

For definition of diversity, some methods consider the training data sets, and others focus on the inner structures of networks. Yang *et al*. [14] proposed the definition based on output curve that provides an effective way for diversity calculation, but the relationship of component network error is not considered.

In this study, the diversity is directly based on the errors of NNs. For convenience of discussion, only single output network and simple average is considered here. m networks are tested with a set of data including n instances. The error of a network is vectorized. All the errors of networks are considered as a high dimension vector, and $\boldsymbol{e}_i = (e_{i1}, e_{i2}, \cdots, e_{in})$ represents the error vector of the ith network. The output error of

Received 2012-05-12, accepted 2012-07-10.
* Supported by the National Natural Science Foundation of China (61074153, 61104131), and the Fundamental Research Funds for Central Universities of China (ZY1111, JD1104).
** To whom correspondence should be addressed. E-mail: xyfancy@163.com

ensemble on the *k*th testing data and the ensemble error vector are respectively

$$e_k = \sum_{i=1}^{m} e_{ik} / m, \quad \boldsymbol{e} = (e_1, e_2, \cdots, e_n) = \sum_{i}^{m} \boldsymbol{e}_i / m \quad (1)$$

With vectorizing, the modulus of the vector is proportional to the output error, and the direction of the vector reflects the error distribution. The direction of the vector is the main factor in judging the diversity.

The error of the *i*th component neural network and the average error are respectively

$$E_i = \left|\boldsymbol{e}_i\right|^2 / n \quad (2)$$

$$\bar{E} = \sum_{i}^{m} E_i / m \quad (3)$$

where $\left|\boldsymbol{e}_i\right|$ is the modulus of $\boldsymbol{e}_i$. It is assumed that there are only two networks in the ensemble, so the output error E of the ensemble is

$$E = |\boldsymbol{e}|^2 = \left|\frac{(\boldsymbol{e}_i + \boldsymbol{e}_j)}{2}\right|^2 \Big/ n = \frac{\left|\boldsymbol{e}_i\right|^2 + \left|\boldsymbol{e}_j\right|^2 + 2\boldsymbol{e}_i \cdot \boldsymbol{e}_j}{4n} \quad (4)$$

Since $2\boldsymbol{e}_i \cdot \boldsymbol{e}_j \leqslant \left|\boldsymbol{e}_i\right|^2 + \left|\boldsymbol{e}_j\right|^2$, $E \leqslant \bar{E}$. The error of ensemble depends on the accuracy of component networks and their diversities. In Eq. (4), $\left|\boldsymbol{e}_i\right|$ and $\left|\boldsymbol{e}_j\right|$ reflect the accuracy of component networks, and the ratio of $\boldsymbol{e}_i \cdot \boldsymbol{e}_j$ and $\left|\boldsymbol{e}_i\right|\left|\boldsymbol{e}_j\right|$ reflects the diversity of the *i*th and *j*th networks. Thus the similarity coefficient α_{ij} and the diversity coefficient β_{ij} can be defined respectively as

$$\alpha_{ij} = \frac{\boldsymbol{e}_i \cdot \boldsymbol{e}_j}{\left|\boldsymbol{e}_i\right|\left|\boldsymbol{e}_j\right|} \quad (5)$$

$$\beta_{ij} = 1 - \alpha_{ij} \quad (6)$$

Parameter β_{ij} is continuous in the interval (0, 2). Eq. (4) indicates that if the accuracy of each network stays the same, the ensemble accuracy increases with β_{ij}. Eq. (6) is also effective in the general form of ensemble output error

$$E = \left|\sum_{i=1}^{m} w_i \boldsymbol{e}_i\right|^2 \Big/ n = \left(\sum_{i=1}^{m} w_i^2 \left|\boldsymbol{e}_i\right|^2 + \sum_{i=1}^{m}\sum_{j=1, j\neq i}^{m} w_i w_j \boldsymbol{e}_i \cdot \boldsymbol{e}_j\right) \Big/ n \quad (7)$$

2.2 Size of NNE

To some extent, the size of ensemble will influence its accuracy. A suitable size for an ensemble can ensure the accuracy and save computation time. For Eq. (7), with similar accuracies of all component networks and the same weights of networks, that is

$$\left|\boldsymbol{e}_i\right|^2 \approx n\bar{E}, \quad w_i = 1/m, \quad \bar{\beta}^{(m)} = \frac{\sum_{i=1}^{m}\sum_{j=1, j\neq i}^{m} \beta_{ij}}{n(n-1)} \quad (8)$$

$$E^{(m)} = \bar{E}/m + (m-1)\bar{E}(1-\bar{\beta}^{(m)})/m = \bar{E} - (m-1)\bar{E}\bar{\beta}^{(m)}/m \quad (9)$$

where $E^{(m)}$ and $\bar{\beta}^{(m)}$ are respectively the output error and average diversity of ensemble that includes *m* component networks. With the assumption that

$$\bar{\beta}^{(m)} \approx \bar{\beta}^{(m+1)} = \bar{\beta} \quad (10)$$

when the number of networks increases from *m* to $m+1$, the reduction of E is

$$\Delta E^{(m)} = E^{(m)} - E^{(m+1)} = \frac{\bar{E}\bar{\beta}}{m(m+1)} \quad (11)$$

Equation (11) indicates that $\Delta E^{(m)}$ will be large when the number of component networks is small, so the ensemble accuracy can be enhanced greatly by increasing the number of networks. For larger *m*, $\Delta E^{(m)}$ is smaller. When $\bar{\beta} < 1$ and $m = 10$,

$$\Delta E^{(10)} = E^{(10)} - E^{(11)} < \frac{\bar{E}\bar{\beta}}{10\times 11} < \frac{\bar{E}}{110} < 1\% \times \bar{E} \quad (12)$$

When the reduction is small enough, it can be ignored. Adding more networks to the ensemble can hardly increase the accuracy when the number of networks is large enough. Selecting ten or so networks will lead to good results.

3 EVSNE METHOD

Selecting several networks with large diversity to form an ensemble may increase the accuracy of the ensemble, but if the diversities among all the trained networks are small, it will be difficult to improve the accuracy by selecting. Thus the premise is to train the networks with large diversities. In this study, based on the error vectorization, EVSNE method is proposed.

In Section 2, the direction of vector reflects the error distribution. When an NN is tested with a set including *n* instances, the error vector is *n*-dimensional. Now the diversity is expressed as the difference of directions of the vectors. If the diversity of error vectors is larger, more components of the error vectors can be offset so that the accuracy of ensemble can be improved greatly. Thus the aim of ensemble is to train the component networks, with which the error vectors can offset others.

In EVSNE, the component networks are back propagation (BP) networks. The networks are trained one after another and the trained networks constitute partial ensemble. Since the direction of the error vector of next network is opposite to that of the previous partial ensemble, a penalty item λ is added to the training error in order to change the current vector direction

and then offset the error of the partial ensemble. For the *i*th network, the training error is altered as follows

$$\boldsymbol{e}_i = \boldsymbol{a}_i - \boldsymbol{d} + \lambda\left(\boldsymbol{a}^{(p)} - \boldsymbol{d}\right) = \boldsymbol{a}_i + \lambda\boldsymbol{a}^{(p)} - (1+\lambda)\boldsymbol{d} \quad (13)$$

For the *k*th training instance of the *i*th network, the training error is

$$e_{ik} = a_{ik} - d_k + \lambda\left(a_k^{(p)} - d_k\right) = a_{ik} + \lambda a_k^{(p)} - (1+\lambda)d_k \quad (14)$$

where a_{ik} is the actual output, e_{ik} is the error, d_k is the desired output, $a_k^{(p)}$ is the actual output of previous partial ensemble, λ is the coefficient of penalty item, $\boldsymbol{a}_i$, $\boldsymbol{e}_i$, $\boldsymbol{d}$, and $\boldsymbol{a}^{(p)}$ are the vector of the varieties.

Then, the square error of BP is altered after adding the penalty item λ,

$$E_{ik} = 1/2e_{ik}^2 = 1/2\left[a_{ik} + \lambda a_k^{(p)} - (1+\lambda)d_k\right]^2 \quad (15)$$

Eq. (15) indicates that the aim of training BP is not only to decrease the error of the current network, but also to offset the error of previous partial ensemble. That is exactly the aim of ensemble. Thus the component networks trained by EVSNE have larger diversities, for the output error vector of each component network can offset that of the former partial ensemble. As a result, the error of ensemble will be decreased.

When the number of networks in the ensemble reaches the set value, the training will stop. According to the principle of selective ensemble, selecting several of the available networks may be better than selecting all of them. With the diversity definition in Eq. (6), the networks with large diversities can be selected easily.

After training all the networks, the diversities of each two networks are calculated and an $m \times m$ matrix is obtained. In Section 2.2, it concludes that selecting ten or so networks will lead to good results. Then traverse the matrix and compare the average diversity of each 10 networks. Finally, 10 networks with larger diversities are selected to form the selective ensemble. Here all the component networks have equal weight (simple average) [3]. The steps of EVSNE are shown in Fig. 1.

4 EXPERIMENTAL

4.1 Experimental data

Three classical data sets for regression are used in the experiment. The data sets are divided into training sets and testing sets. And the validation sets used for selecting component networks are bootstrap sampled from their training sets, which is normally half the size of training sets. The parameters of the data sets are shown in Table 1. The housing and concrete data sets are from UCI machine learning repository, and the freidman#1 data sets are generated from the equations as follows.

Freidman:

$$t = 10\sin(\pi x_1 x_2) + 20(x_3 - 0.5)^2 + 10x_4 + 5x_5 + \varepsilon$$

$$x_i \sim U[0,1] \quad \varepsilon \sim N[0,1]$$

where $U[0, x]$ stands for a uniform distribution whose interval is 0 to x.

Table 1 Data sets

Data sets	Variable number (input/output)	Total sample (training/testing)
housing	13/1	506 (380/126)
concrete	8/1	1030 (800/230)
Freidman#1	5/1	1400 (800/600)

4.2 Experimental results

In this experiment, BP network is used as the component network. The number of hidden layer node is 9, the learning efficiency and the momentum item coefficient are 0.5 and 0.3, respectively, and each network is trained for 2000 times.

Firstly, 25 BP neural networks are trained by EVSNE. Secondly, some instances are bootstrap sampled from the training set as the validation set, which is half the size of training set. Thirdly, the error vector and the diversity matrix are calculated. Finally, 10 networks with largest diversity are selected to make up the ensemble. The generalization error is measured as relative error

$$E = \sum_{k=1}^{n} \frac{|a_k - d_k|}{nd_k} \times 100\% \quad (16)$$

where a_k is the actual output, d_k is the desired output, and n is the number of instances. In order to exclude the influence of occasionality, three experiments are taken independently.

4.2.1 *Selection of penalty item*

The value of penalty item λ is crucial to EVSNE. If it is too small, the offsetting effect will be weak, so

$$\begin{cases} 1 \text{ Train the 1st network normally} \\ 2 \text{ for } t = 1 \text{ to } m \begin{cases} \text{Calculate the output error of partial ensemble} \\ \text{Train a new network by the improved function [Eq. (15)] and add it to the partial ensemble} \end{cases} \\ 3 \text{ Calculate the diversity of each two networks to obtain an } m \times m \text{ matrix} \\ 4 \text{ Traverse the matrix and select several networks with large diversities to constitute the ensemble} \end{cases}$$

Figure 1 Steps of EVSNE method

Table 2 Testing results of the three methods on standard data sets

	Housing		Concrete		Freidman#1	
	Generalization error/%	Standard deviation	Generalization error/%	Standard deviation	Generalization error/%	Standard deviation
Bagging	12.91	0.144	16.57	0.255	10.22	0.122
DWSEN	11.25	0.113	13.25	0.197	9.44	0.092
EVSNE	10.31	0.137	12.42	0.189	7.38	0.098

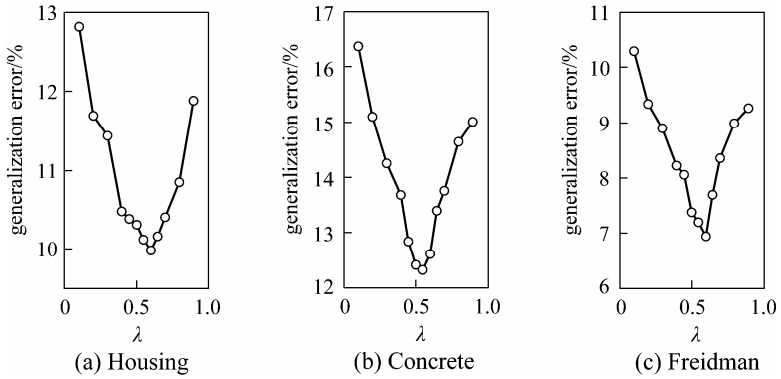

Figure 2 Effect of penalty item λ on three data sets

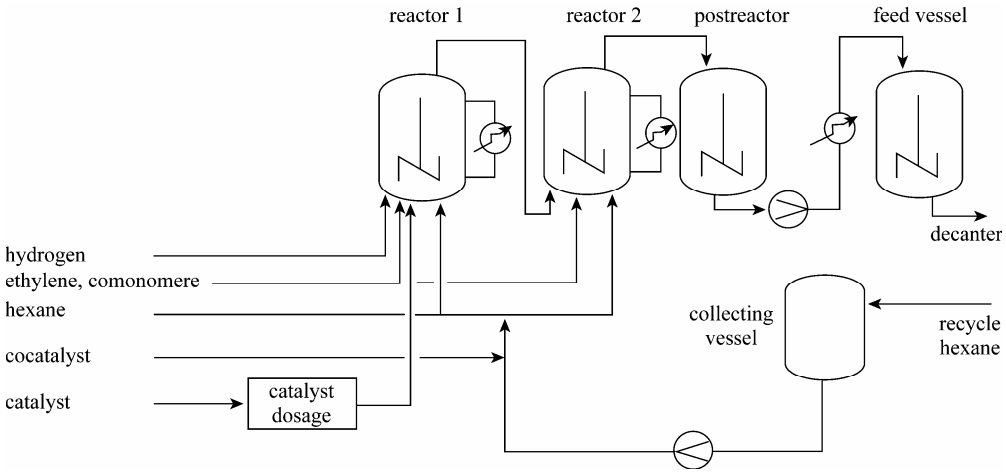

Figure 3 Flowchart of HDPE cascade reaction

that the diversities are small. If it is too large, the errors of the component networks will be large, resulting in large ensemble error. The relationship between the normalized generalization mean square error of SEL-NNE and the penalty item λ is shown in Fig. 2, in which λ varies from 0.1 to 0.9. The overall trends of generalization error for different data sets are nearly the same with the increase of penalty item λ. When the penalty item λ varies between 0.5 and 0.6, EVSNE method performs best. In the following sections, all the experiments are carried out for λ of 0.5.

4.2.2 *Generalization result*

In this section, the generalization result of Bagging, DWSEN (diversity measurement using weights based selective ensemble) [10] and EVSNE are compared. Table 2 shows the generalization relative error of the housing, concrete and freidman#1 data sets. EVSNE has better generalization ability than Bagging or DWSEN. It is proved that the networks trained by EVSNE have larger diversities than by Bagging or DWSEN. The accuracies of ensemble are all improved after selecting, which proves the feasibility of the selecting method.

5 APPLICATION EXAMPLE

5.1 HDPE process

Actual chemical processes are complex, so neural network ensemble is usually used. HDPE (high-density polyethylene) cascade reaction [15–18] is a polymerization process from high-purity ethylene monomer to high-density polyethylene under the condition of low-pressure and hexane slurry. It is composed of two reactors connected by flash tank in series, post reactor,

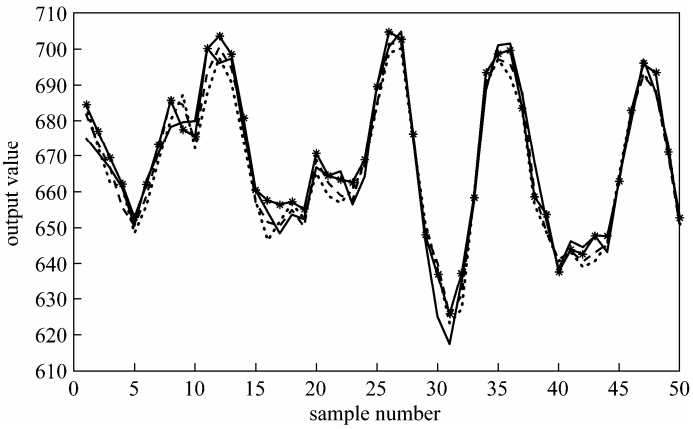

Figure 4 Generalization output curve on HDPE data set
····· Bagging; – – – DWSEN; —✳— EVSNE; —— desired value

and recycle unit, the flowchart of which is shown in Fig. 3. The hexane solvent is added into reactors 1 and 2 under certain temperature and pressure. Hydrogen and high-purity ethylene are mixed with dehydration and impurity removal, and then are collected into reactors. Catalyst and co-catalyst are stirred at certain ratio, and then injected to the reactors 1 and 2 separately. With regulating the feed flow, controlling the ratio of hydrogen to ethylene, and using the effect of catalyst, slurry polymerization is carried out in the hexane solvent system. The heat of polymerization is removed in the form of the latent heat of vaporization of hexane solvent, cooling water in the jacket of reactors, and external reflux of slurry.

In reactor 1, the ratio of hydrogen and ethylene is large so that the melt index of the product is high. The product from reactor 1 is flash-evaporated to recycle some micro-molecule hydrocarbon and is collected into reactor 2 for further polymerization. In an actual process, the melt index of reactor 1 is an important quality index. From the technical mechanism and production experiences, the melt index of reactor 1 mainly depends on seven variables, *i.e.*, the flow rates of catalyst, hydrogen and ethylene, reaction temperature and pressure, ethylene pressure, and pressure ratio of hydrogen and ethylene. Thereby, exploring the relation between the melt index and process variables is crucial to provide effective production guide.

5.2 Experimental result

For the process modeling, the seven variables affecting the melt index significantly are selected as the input variables, and the melt index is taken as the output variable. 1200 instances from the data sets are collected, in which 700 of data are used as the training set and the remaining is used as the generalization set. For a chemical process the relative error is a crucial factor [19, 20], and the relative error and their standard deviation are shown in Table 3. EVSNE can reduce the generalization error efficiently. Its relative error is smaller than that of Bagging or DWSEN, so that EVSNE gives the melt index more exactly. The generalization curves of Bagging, DWSEN and EVSNE are shown in Fig. 4, in which only the first 50 samples are used for clearness, and all the instances are denormalized from (0, 1) to the real value. The generalization curve of EVSNE is more close to the actual value than that of Bagging or DWSEN, so EVSNE can provide efficient guide for chemical production.

Table 3 Testing results on HDPE data set

	Generalization error/%	Standard deviation
Bagging	0.7352	0.0215
DWSEN	0.5991	0.0187
EVSNE	0.5103	0.0359

In order to see the effect of ensemble size, 5−14 networks are selected separately from ALL-NNE to form the SEL-NNE. The relationship between the number of networks in the SEL-NNE and generalization error is shown in Fig. 5. When the number of networks in SEL-NNE is larger than ten, the variation of generalization error is small. Thus it is suitable to select ten or so networks to form SEL-NNE.

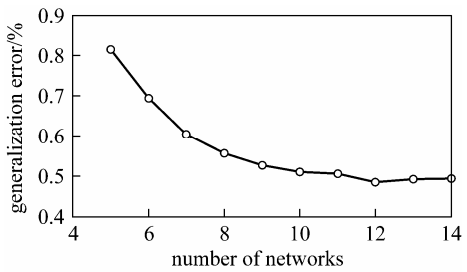

Figure 5 Generalization error on HDPE data set

6 CONCLUSIONS

Actual chemical processes are complex, so neural

network ensemble methods are usually used for modeling. For the selective neural network ensemble, some problems have to be solved, *e.g.*, there is no unified definition among networks and it is difficult to select a group of networks with large diversity if the diversities of available networks are all small. In this study, EVSNE is proposed, in which the diversity of networks is defined by error vectorization, the number of networks in the ensemble is analyzed, the networks are trained orderly, and the error of the current network can offset the former partial ensemble. The component networks trained by EVSNE have large diversities, since the error vector of each network can offset that of each other. From experiments and comparisons over standard data sets and the data sets from HDPE cascade reaction process, the diversities of networks trained by EVSNE are large, so that the generalization capability is enhanced obviously with higher accuracy.

REFERENCES

1 Hansen, L.K., Salamon, P., "Neural network ensembles", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **12** (10), 993−1001 (1990).

2 Sollich, P., Krogh, A., "Learning with ensembles: How over-fitting can be useful", *Advances in Neural Information Processing Systems*, **8**, 190−196 (1996).

3 Zhou, Z.H., Chen, S.F., "Neural network ensemble", *Chinese J. Computers*, **25** (1), 1−8 (2002). (in Chinese)

4 Akhand, M.A.H., MonirulIslam, Md., Murase, K., "Progressive interactive training: A sequential neural network ensemble learning method", *Neurocomputing*, **73**, 260−273 (2009).

5 Li, K., Huang, H.K., "Study of a neural network ensemble algorithm for small data sets", *Journal of Computer Research & Development*, **43** (7), 1161−1166 (2006). (in Chinese)

6 Breiman, L., "Bagging predictors", *Machine Learning*, **24** (2), 123−140 (1996).

7 Schapire, R.E., "The strength of weak learnability", *Machine Learning*, **5** (2), 197−227 (1990).

8 Zhou, Z.H., Wu, J.X., Tang, W., "Ensembling neural networks: Many could be better than all", *Artificial Intelligence*, **137** (122), 239−263 (2002).

9 Wu, J.X., Zhou, Z.H., Shen, X.H., Chen, Z.Q., "A selective constructing approach to neural network ensemble", *Journal of Computer Research & Development*, **37** (9), 1039−1044 (2000). (in Chinese)

10 Zhu, Q.X., Meng, Q.H., "A new selective neural network ensemble method and its application in purified terephthalic acid solvent system", *CIESC Journal*, **60** (10), 2511−2516 (2009). (in Chinese)

11 Fu, Q., Hu, S.X., Zhao, S.Y., "PSO-based approach for neural network ensemble", *Journal of Zhejiang University* (*Engineering Science*), **38** (12), 1596−1599 (2004). (in Chinese)

12 Fu, Q., Hu, S.X., Zhao, S.Y., "Clustering-based selective neural network ensemble", *Zhejiang Univ SCI*, **6A** (5), 287−392 (2005). (in Chinese)

13 Li, K., Huang, H.K., "A selective approach to neural network ensemble based on clustering technology", *Journal of Computer Research & Development*, **42** (4), 594−598 (2005). (in Chinese)

14 Yang, Y., Xu, Y., Zhu, Q.X., "A neural network ensemble method with new definition of diversity based on output error curve", In: 2010 IEEE International Conference on Intelligent Computing and Intelligent Systems(ICISS), Zhang, Z., Zhou, X.Y., Liu, P., eds., IEEE Press, China, 587−590 (2010).

15 Xu, Y., Zhu, Q.X., "Multi-layer moving-window hierarchical neural network for modeling of high-density polyethylene cascade reaction process", In: 2010 11th International Conference on Control Automation Robotics & Vision(ICARCV), IEEE Computer Society, USA, 1684−1687 (2010).

16 Zhu, X.H., Guo, Z.F., "Ethylene polymerization using improved polyethylene catalyst", *Chin. J. Chem. Eng.*, **19** (1), 52−56 (2011).

17 Guo, Z.F., Chen, W., Zhou, J.L., Yang, H.X. "Novel high performance Ziegler-Natta catalyst for ethylene slurry polymerization", *Chin. J. Chem. Eng.*, **17** (3), 530−534 (2009)

18 Wang, J., Yang, Y., "Study on optimal strategy of grade transition in chemical fluidized bed gas-phase polyethylene production process", *Chin. J. Chem. Eng.*, **11** (1), 1−8 (2003).

19 Zhu, Q.X., Li, C.F., "Dimensionality reduction with input training neural network and its application in chemical process modelling", *Chin. J. Chem. Eng.*, **14** (5), 597−603 (2006).

20 Zhang, Y., Liang, X.X., Yang, P., Chen, Z.Q., Yuan, Z.Z., "Modeling and control of nonlinear discrete-time systems based on compound neural networks", *Chin. J. Chem. Eng.*, **17** (3), 454−459 (2009).