SOFT-SENSING OF THE DRY POINT OF BENZENE USING PCA AND DRBFN

Yuqing Chang and Fuli Wang

Dept. of Information Science and Engineering, Northeastern University
Shenyang, Liaoning Province, China

Furong Gao
Department of Chemical Engineering
Hong Kong University of Science & Technology
Clear Water Bay, Kowloon, Hong Kong

Abstract: Measurements of temperatures and flows and pressures are used to estimate the dry point of the product for the distillation column. The Problem is characterized by the model complication and the strong colinearity between the measurements. In this article, the distributed RBF neural network (DRBFN) and principal component analysis (PCA) are used to develop the soft sensor (PCA-DRBFN model), and PCA is also used for data compressing and validation. Another two models are used to compare the performance with the proposed soft sensor.

Key words: Soft Sensor, Estimation, PCA, DRBFN, Dry Point, Distillation Column

1. INTRODUCTION

A major problem in the control of product quality in chemical process is the uneasy measuring of the quality variables on-line. Although related product quality parameters (such as product composition) can be obtained by laboratory analysis off-line, this brings large measurement delays. This paper addresses the development of a soft sensor model to achieve the estimation of the uneasy measured quality parameter. The application chosen here is the use of temperature, flow and pressure parameters to estimate the product dry point.

There are many methods of developing soft sensor models and neural network is one of them being used widely because of its excellent properties (Bhat and McAvoy 1990). RBF neural network is the often-used net.

Usually, it is necessary to collect large amount of process data in order to accurately developing the soft sensor model. In this case, using one network to build model will bring a problem of long learning time. Distributed RBF network (DRBFN), which learns all the initial data using multi-nets can deal with this problem properly. However, there is usually strong colinearity among the multi-dimension variables in chemical process, and this will lead to ill-condition model, long learning time and huge model structure. Principal components analysis (PCA) technology can compress the multidimensional collinear variables into lower dimension and guarantee the least loss of data information, so principal component regression (PCR) can be used to develop the estimation model and avoid the shortcoming from colinearity variables. However, PCR are only fit to linear regression, so this method will bring bad estimation result for the complicated nonlinear chemical process (such as distillation column).

This article proposed a new soft sensor model using PCA and DRBFN technoledges. The proposed model is of the specialties of better estimation quality and simplified structure compared with the PCR and DRBFN model.
Although it is based on a particular distillation column example, the treatment in this article is rather general.

2. DRBFN SOFT SENSOR

The objective is to obtain the best prediction $\hat{y}$ of the primary variable (product dry point in our application) using all available information. The estimation model (soft sensor model) may be written

$$\hat{y} = f(X) \tag{1}$$

where, X includes all measured secondary variables.

The structure of DRBFN soft sensor is shown in Figure 1 (Wang and Shao, 1998).

In Figure 1, RBFi (i=1,2…n) is the sub RBF network. The fuzzy classifying unit is used to classify the initial learning data into n classes using Rival Penalized Compete Learning algorithm, and the RBFi net matches the data relation of the i*th* class. The combination of all the RBF subnet is realized by the membership degree $\mu = [\mu_1, \mu_2, \cdots, \mu_n]$. μ_i is achieved by the fuzzy classifying unit in Figure 1, it can be written

$$\begin{cases} \eta_i = 1, \eta_j = 0, & if \;\; d_i = 0, j = 1, \cdots N, j \neq i \\ \eta_i = \left(1 \big/ d_i \right) \Big/ \left(\sum\limits_{i=1}^{N} 1 \big/ d_i \right), & otherwise \end{cases} \tag{2}$$

$$\mu_i = \sum_{j}^{N_i} \eta_j \text{ 。} \tag{3}$$

where, $d_i = \| X - X_i \|^2$ is the Eurlar distance between the input data X and the initial sample data $X_i (i = 1, 2, \cdots, N)$, N and N_i are the total number of the sample data and the number of i*th* class sample data respectively.

The output of DRBFN can be expressed

$$Y = \sum_{i=1}^{n} \mu_i f_i(X) \tag{4}$$

Where, f_i is the output of i*th* RBF subnet, n is the number of all the RBF subnets.

The advantages of DRBFN soft sensor are that it can approximate any continuous nonlinear functions and avoid the long learning time from the large number of the sample data. However, if X is the multidimensional variable and has significant collinearity, the input of each RBF subnet will be of serious redundancy, and this will lead to ill-conditioned model, long learning time and complicated subnet structure.

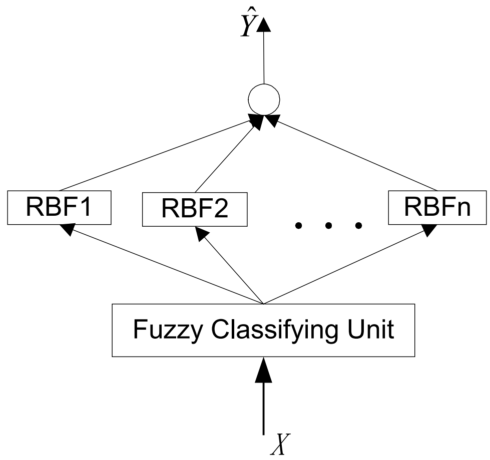


Fig.1. Soft Sensor Structure Based on the DRBF network

3. PCA-DRBFN SOFT SENSOR

3.1 DATA COMPRESSION AND VALIDATION

PCA is an extremely powerful method for data compression, and has been successfully used to a wide variety of different applications. It is at its best when applied to problems featuring both high dimensionality and a large degree of collinearity. PCA breaks data matrices $\boldsymbol{X}(N \times m)$ down into a series of abstract latent variables or principal components. Its model is given by

$$\boldsymbol{X} = \boldsymbol{T}\boldsymbol{P}^{\boldsymbol{T}} + \boldsymbol{E} = \sum_{i=1}^{l} t_l p_l^T + \boldsymbol{E} \tag{5}$$

where, $\boldsymbol{T} = [t_1, t_2 \cdots t_l]_{N \times l}$ is the score vectors, $\boldsymbol{P} = [p_1, p_2 \cdots p_l]_{m \times l}$ is the loading vectors and $\boldsymbol{E}$ is the residuals of the $\boldsymbol{X}$ blocks. In this paper, the PCA approach is adopted to compression the original process data. Based on the above PCA method, the PCR model can be given by

$$\hat{Y} = K_{PCR} X$$

and $$K_{PCR} = \boldsymbol{Y}^T \boldsymbol{T}(\boldsymbol{T}^T \boldsymbol{T})^{-1} \boldsymbol{P}^T$$

where, Y and $\hat{Y}$ are the measurement and prediction of the primary variable respectively.

In Figure 2, the PCA demapping unit is used to regress the original variables $\hat{X}$ by the score vectors and loading vectors. This model can be expressed as

$$\hat{\boldsymbol{X}} = \boldsymbol{T}\boldsymbol{P}^T \tag{6}$$

The squared prediction error (SPE) for $\boldsymbol{x}^j$ is

$$SPE(\boldsymbol{x}^j) = \sum_{i=1}^{m} SPE_i = \sum_{i=1}^{m} (x_i^j - \hat{x}_i^j)^2 \tag{7}$$

where, $\boldsymbol{x}^j = [x_1^j, x_2^j \cdots x_m^j]$ is the sample value of the *j*th sample period, and *m* is the dimension of original variables. If one sensor fails which breaks the normal correlation, the SPE will increase significantly. Jackson and Mudholkar developed a test for SPE known as the Q-statistic. This test suggests the existence of an failure sensor when

$$SPE > Q_\alpha \tag{8}$$

where

$$Q_\alpha = \theta_1 \left[\frac{C_\alpha \sqrt{2\theta_2 h_0^2}}{\theta_1} + 1 + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} \right]^{\frac{1}{h_0}} \tag{9}$$

$$\theta_i = \sum_{j=l+1}^{m} (\lambda_j)^i \quad i = 1,2,3 \tag{10}$$

$$h_0 = 1 - \frac{2\theta_1 \theta_3}{3\theta_2^2} \tag{11}$$

and C_α is the confidence limit for the $1-\alpha$ percentile in a normally distributed.
Defines the SPE contribution β_i as

$$\beta_i = SPE_i / SPE \quad i = 1,2 \cdots m \tag{12}$$

The data x_i^j is fault when

$$\beta_i > \delta \tag{13}$$

where, δ is a given value. From (6), we have

$$\hat{\boldsymbol{X}} = \boldsymbol{T}\boldsymbol{P}^T = \boldsymbol{X}\boldsymbol{P}\boldsymbol{P}^T \tag{14}$$

and

$$\boldsymbol{P}\boldsymbol{P}^T = [\overline{\boldsymbol{w}}_1, \overline{\boldsymbol{w}}_2 \cdots \overline{\boldsymbol{w}}_m] \tag{15}$$

The validation of fault data x_i^j can be expressed as

$$\hat{x}_i^j = \boldsymbol{x}^j \overline{\boldsymbol{w}}_i \tag{16}$$

In this paper, the above PCA method will be used to realize the data compression and validation.

3.2 PCA-DRBFN SOFT SENSOR

The structure of the soft sensor model based on PCA-DRBFN is shown in Figure 2, where, X and $\hat{Y}$ are the secondary variable vector and primary variable vector respectively.

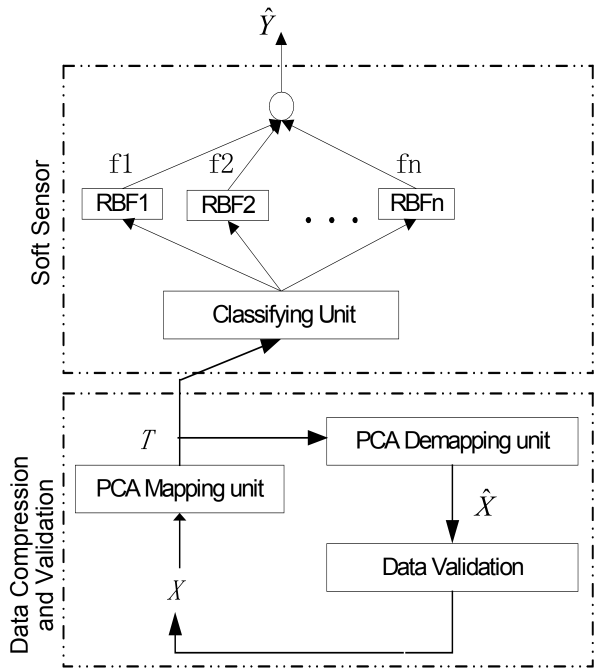


Fig. 2. Soft Sensor Structure Based on the PCA-DRBFN

In Figure 2, the Data compression and validation unit is used to compress the original higher dimensional secondary variables X into principal component variables T, and to validate the fault process data so that the process information used by soft sensor is compact and available. The relationship of T and X can be expressed by equation 5. In PCA-DRBFN model, the input of each subnet is T instead of X. By this means, the input of each subnet will be decreased from *m* to *l* (m>>*l*) if the secondary variables are collinear. So the RBF subnet structure of PCA-DRBFN can be significantly simplified by PCA method, and the learning speed of the net can also be improved.

The output of the PCA-DRBFN can be written as

$$Y = \sum_{i=1}^{n} \mu_i f_i(T) = \sum_{i=1}^{n} \mu_i f_i(XP) \qquad (17)$$

where, μ_i and f_i are designed in Section 2.

4. DRY POINT PREDICTION SIMULATION

In this paper, a benzene-distilled process is adopted as a simulation sample to test the validation of the proposed model. This column consists of 35 trays and its diameter is 1.6m. A reboiler is used to heat the raw material. A water-cooled condenser is placed in the top of the column and a small accumulator tank is used to deposit the condensate. The structure of distillation column is shown in Figure 3, and the main process variables have been marked on the sides.

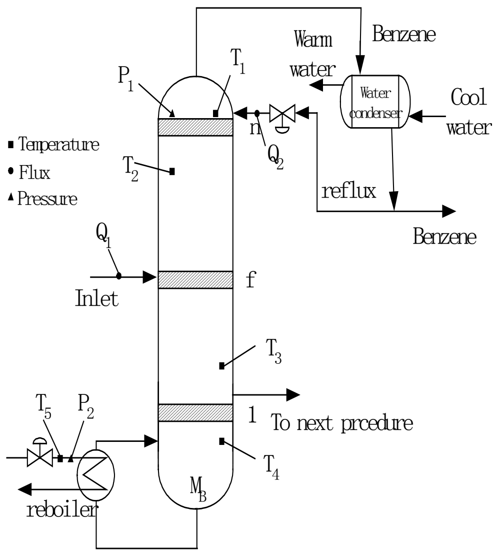


Fig.3. The Flow Chart of the Distillation Column of the Benzene

Tabel 1. Secondary variables

Index		Name of the variables
1	P_1	tower top pressure (atm)
2	T_1	tower top temperature(℃)
3	T_2	tray 28 temperature(℃)
4	T_3	tray 4 temperature(℃)
5	T_4	tower bottom temperature(℃)
6	P_2	steam pressure (atm)
7	T_5	steam temperature(℃)
8	Q_1	inlet flux(m^3/hour)
9	Q_2	reflux(m^3/hour)

The top product of this column is pure benzene. The dry point, which is achieved by sample analysis offline with a long measuring delay, is used to evaluate the quality of the product. The soft sensor based on the PCA-DRBFN is used to obtain the prediction of product dry point.

The variables that affect the dry point of the product are listed in table 1. The soft sensor model of the dry point y can be expressed as

$$y = f(P_1, T_1, T_2, T_3, T_4, P_2, T_5, Q_1, Q_2) \quad (18)$$

4.1 MODEL PREDICTION

Two hundred data points were collected from a distillation process. The variable to be predicted is the product dry point sampled by laboratory analysis. The data were collected so as to achieve the soft sensor based on the process information. In the simulation, 150 data points are used for building the PCA-DRBFN soft sensor and 50 points are used to test the generalization property of the model. After the principal component analysis for the 150 data, the contribution percent of each PC is shown in Table 2.

From Table 2, the former 4 PCs' cumulative contribution is 87.23%, so these 4 PCs can describe the information of process and filter the redundancy (Dunteman, 1989). The variables dimension is decreased from 9 to 4 after PCA. It means the net structure will be simplified significantly and the learning time of each RBFi will also be decreased.

After principal component analysis, let the compressed data input into the distributed RBF to obtain the soft sensor model. At the same time, we use the same initial sample data to develop the DRBF network soft sensor and PCR soft sensor.

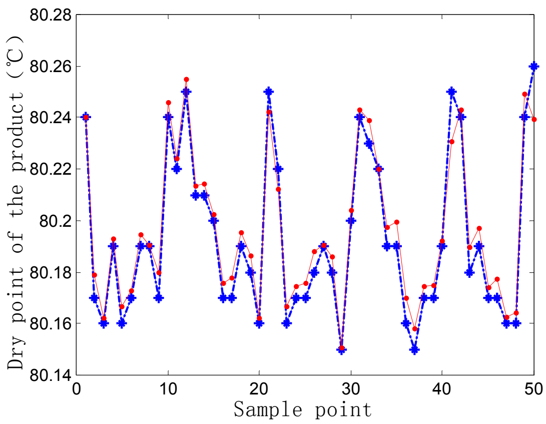
50 test data are used to test the above two soft sensor. Figure 4 shows prediction results of the PCA-DRBFN soft sensor. It shows that the proposed soft sensor model can achieve the prediction value of the product dry point with a considerable precision. The estimation errors of the above three models are showed in Figure 5.

From Figure 5, we can see that the estimation quality of the DRBFN and the PCA-DRBFN is similar, but the estimation result of the PCR is deteriorated because of the higher nonlinearity of the process. Although the estimation quality of the DRBFN and the PCA-DRBFN is similar, the structure size of them is different. The biggest subnet size of DRBFN is 9×21×1(that means it has 9 nodes in input layer, 21 nodes in hider layer and 1 node in output layer), the smallest one is 9×15×1, and the biggest subnet size of PCA-DRBFN is

4×11×1, the smallest one is 4×7×1, so the structure of PCA-DRBFN is simplified. By simulation, we also find that the learning time of PCA-DRBFN is shorter than that of DRBFN.

Table 2 PC contribution percent

	Principal Component								
	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9
Latent root	2.71	2.40	1.58	0.66	0.45	0.26	0.19	0.13	0.05
Contribution percent (%)	32.2	28.5	18.7	7.83	5.33	3.08	2.24	1.54	0.58
Cumulative Contribution percent (%)	32.2	60.7	79.4	87.23	92.56	95.64	97.88	99.42	100



* prediction value • measured value

Fig.4. The Prediction of the Dry Point Based on the PCA-DRBF Soft Sensor

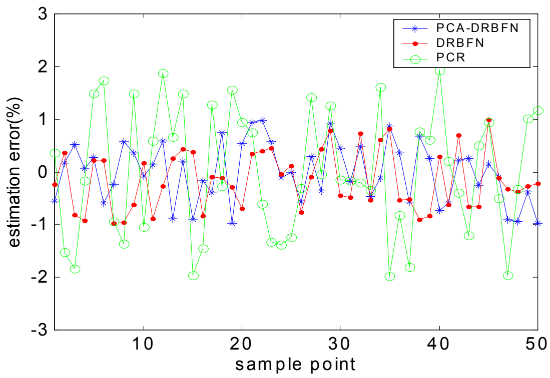


Fig.5. The Estimation Error of the Soft Sensor Model

4.2 DATA VALIDATION

Two hundred data points were collected with a bias fault introduced in the inlet flux so as to test the property of the data validation model. Based on the obtained PCA model, the SPE value of the testing data can be calculated on-line. Figure 6 shows the SPE of the data is out of the control limit after 90th sample, and Figure 7 shows the SPE contribution of the 95th sample point. From Figure 7, inlet flux Q_1 need to be reconstructed, and the SPE after data validation is shown in Figure 8. The result shows SPE returns to the normal range after the faulty data being reconstructed.

5. CONCLUSIONS

In this paper, a method of building a soft sensor model is proposed, and PCA method is used to compress the higher dimensional secondary variables, so that the soft sensor has a compact model structure. The simulation shows that the proposed soft sensor based on PCA-DRBFN can predict the uneasy measured quality variable accurately.

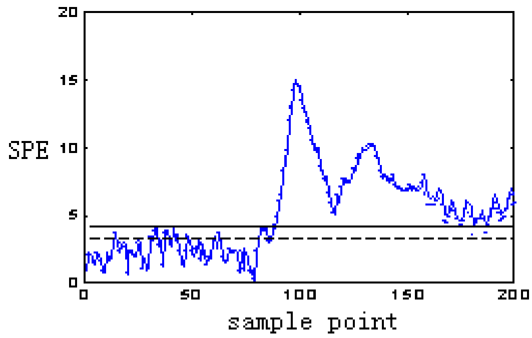


Fig.6 SPE test for faulty data

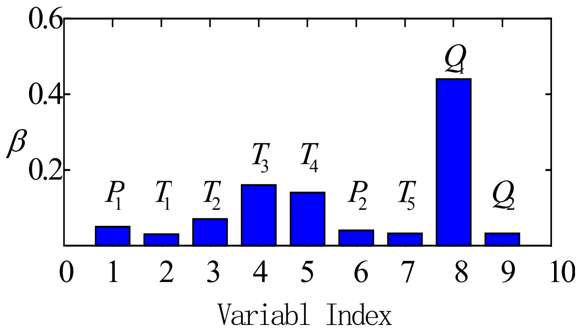


Fig.7 SPE contribution chart

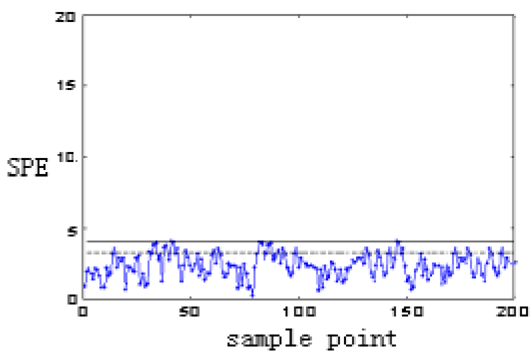


Fig.8 SPE test after data reconstruction

REFERENCES

Bhat, N. V. and T., J. McAvoy, , Use of neural nets for dynamical modeling and control of chemical process systems, Computers and Chemical Engineering, 1990, 14, 573-583

Thor Mejdell and Sigurd Skogestad, Output Estimation Using Multiple Secondary Measurements: High-Purity Distillation, AIChE J, 1993, Vol.39, No.10, 1641-1653

Xudong Wang and Huihe Shao. Distributed RBF neural network and its application in soft sensor, Control theory and application, 1998, Vol.15, No4, 558-563

Huiwen Wang, Partial Least-Squares Regression-Method and Application, National Defence Industry Publishing Company, 1999

George H. Dunteman, “Principal component analysis”, SAGE publication, 1989

Wold S., “Cross-validatory estimation of the number of components in factor and principal component models”, Technometrics, 1978, 20, 397-405

George H. Dunteman, “Principal component analysis”, SAGE publication, 1989.

Jackson, J.E., and Mudholkar, G. Control procedures for residuals associated with principal component analysis. Technometrics 1979,21,341-349

S.Joe Qin, Hongyu Yue and Ricardo Dunia, A Self-Validating Inferential Sensor for Emission Monitoring, Proceedings of the American Control Conference, 1997,473-477