Proceedings of the Second International Conference on Machine Learning and Cybernetics, Xi'an, 2-5 November 2003

A SOFT SENSOR BASED ON NONLINEAR PRINCIPAL COMPONENT ANALYSIS

YU-HONG ZHAO

Institute of Systems Engineering, Zhejiang University, Hangzhou, 310027, P.R.China
E-MAIL: yhzhao@iipc.zju.edu.cn

Abstract:

An accurate on-line measurement of quality variables is essential for the successful monitoring and control tasks in chemical process operations. A soft sensor is developed based on nonlinear principal analysis (PCA), due to the ability of capturing the linear and nonlinear features of the data. The proposed method is applied to an industrial crude oil atmospheric distillation tower and is illustrated by comparisons with other familiar methods. The results have shown that the proposed method gives a better or equal performance over the conventional PCA method and neural networks method.

Keywords:

Soft sensor; Nonlinear PCA; Neural network;

1 *Introduction*

For successful monitoring and control of chemical plants, there are important quality variables that are difficult to measure on-line, due to limitations such as cost, reliability, and long dead time. These measurement limitations may cause important problems such as product loss, energy loss, toxic byproduct generation, and safety problem. A soft sensor, an inferential model, can estimate the qualities of interest on-line using other available on-line measurements such as temperatures and pressures. The soft sensor can be derived from the first principal model when the model offers the sufficient accuracy within the reasonable computation time. However, there are cases when the first principal model is not available, or sometimes it takes too much time to compute. As a result, empirical models are the most popular ones to develop soft sensors. Empirical models are usually obtained based on various modeling techniques such as multivariate statistics and artificial neural networks[1].

Multivariate statistical methods based on linear projection, especially principal component analysis (PCA), have attracted wide interests as robust methods for constructing empirical models[2-4], particularly when there are high dimensionality and collinearities in the data.

Although these linear projection methods can handle high dimensionality and collinearity, their major restriction is that only linear information is extracted from the data. Since many chemical processes are nonlinear in nature, it is desirable to have a robust method which can model any nonlinearity. Artificial neural networks have found many applications in nonlinear regression modeling at present[5,6]. Among all types of neural networks, multilayer feed-forward networks have primarily been applied in the process modeling because of their excellent ability to approximate nonlinear functions.

A new nonlinear method, which is conceptually simple and easy to use, is proposed in this paper. A nonlinear principal component analysis model is established through a feed-forward neural network. Then the linear relationship between the scores and estimated variable (primary variable) is gained by robust linear regression based on M-estimation. Since the nonlinear PCA captures the nonlinear features in the measurements (secondary variable), the derived soft sensor provides a perfect image between the primary variables and the secondary variables.

In the following a description of principal component analysis and nonlinear principal component analysis is proposed in section 2. The construction of the soft sensor based on the nonlinear PCA model is given in section 3. The application of a crude oil atmospheric distillation tower and the comparison with conventional PCA method and neural network method are provided in section 4. The conclusions are drawn and the further research is indicated in the last section.

2 Nonlinear Principal Component Analysis

Principal component analysis (PCA) is a multivariate statistical technique that is capable of treating high dimensional, noisy and correlated data by projecting it onto a lower dimensional subspace that explains the most pertaining features of the system[7]. As the PCA is described so often in many research papers, only an overview is given in this paper.

Let $X \in R^{N \times m}$ represent the data matrix (N

0-7803-7865-2/03/$17.00 ©2003 IEEE

designates the number of observations, and m denotes the number of variables). After mean centering and scaling, a positive semi-definite covariance matrix of the data set X can be decomposed to find eigenvectors p, associated with m eigenvalues. The first principal component, t_1, which is the linear combination of m variables, explains the greatest amount of variability.

$$t_1 = Xp_1 \tag{1}$$

and the second principal component explains the next greatest amount of variability subject to the orthogonality condition: $p_1 \perp p_2$

$$t_2 = E_1 p_2 \tag{2}$$

where E_1 is the residual matrix given by

$$E_1 = X - t_1 p_1^T \tag{3}$$

One can continue to find m principal components, but then, there will not be any dimensionality reduction. Due to high degree of correlation among variables, one often finds k principal components ($k \ll m$) sufficient to explain most of the relevant information contained in the data matrix X. Algebraically, the X matrix can be reconstructed by

$$X = t_1 a_1^T + t_2 a_2^T + \cdots + t_k a_k^T + E = TP^T + E \tag{4}$$

where $T = [t_1, t_2, \cdots, t_k]$ is defined to be the matrix of principal component scores, $P = [p_1, p_2, \cdots, p_k]$ is the matrix of principal component loadings and E is the residual matrix. Without the residual matrix E, which contains the sole redundant information, Eq. (4) is referred to as the PCA model that encompasses the major variation in the data.

However, PCA are able to capture only the linear features of the data. To address the nonlinearity in the identity mapping of multivariate data, an extension of PCA, nonlinear PCA is proposed in previous studies. A principal curves methodology proposed to provide a nonlinear summary of an m-dimensional data set in the first instance. However, this approach is non-parametric and cannot be used for continuous mapping of new data. The parameterization problem is resolved with the use of auto associative neural network, and it is referred to as the nonlinear principal component analysis (NLPCA). NLPCA uses a four-layer feed-forward neural network (FFNN) to carry out the identity mapping. The method uncovers both linear and nonlinear correlations independent of the structure of nonlinearity present in the data. The neural network architecture of NLPCA is given in Fig. 1. NLPCA operates by training a FFNN to perform an identity mapping where network inputs are reconstructed at the output layer. The bottleneck layer contains compact representation of the input data and its dimension is lower both from input and output layers.

In the NLPCA, mapping onto the lower dimension is generalized such that the nonlinear features are captured. In

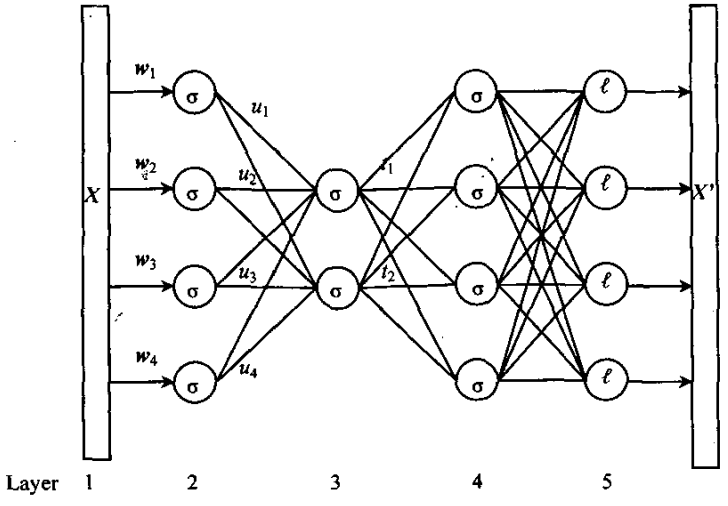

Figure 1. Neural network architecture of the nonlinear component analysis
Note: Linear activation functions are denoted by ℓ and sigmoidal functions by σ

analogy to Eq. (1), the following form of the mapping is sought:

$$T = J(X) \tag{5}$$

where J is a nonlinear vector function composed of k individual nonlinear functions, and $T = T(t_1, t_2, \cdots, t_k)$. By preserving the original dimension of the data, inverse transformation is performed in analogous to Eq. (4) except using a nonlinear vector function L.

$$X = L(T + E) \tag{6}$$

In this case, the objective function to be minimized is as follows:

$$\min \sum_{i=1}^{N} \left\| x_i - L(t(x_i)) \right\|^2 \tag{7}$$

It has been shown that any nonlinear function could be approximated to an arbitrary degree of precision using the following structure:

$$v_k = \sum_{j=1}^{N_2} w_{ik2} \sigma \left(\sum_{i=1}^{N_1} w_{ij1} u_i + \xi_{j1} \right) \tag{8}$$

where $\sigma(x)$ can be any continuous and monotonically increasing function. w_i is the weights assigned to each neuron, u_i is inputs from previous network layer, and ξ_i is biases. N_1 and N_2 are the numbers of neurons in previous and current layers respectively. Here in order to obtain the nonlinear principal components, the function

$$\sigma(x) = \frac{2}{1 + e^{-2x}} - 1 \tag{9}$$

and back-propagation (BP) training algorithm are adopted to train the nonlinear PCA network.

3 Soft Sensor Based on Nonlinear PCA

3.1 Variable Selection

An industrial database usually provides all the variables that can be recorded. However, it is not necessarily true that all recorded variables are relevant to the process variable to be predicted. Instead, it is often the case that when some of the irrelevant variables are deleted from the database, the modeling results can be improved; if the irrelevant variables are kept in the model, they play a role in the noise which can potentially deteriorate the modeling results. Therefore, it is imperative to select a subset of process variables that are truly relevant to the predicted variables.

Prior knowledge is used to screen out totally irrelevant variables first in this paper. Then cross-validation method is conducted to determine whether a variable is truly contributing to the output variable. The prediction error on the test set is calculated before and after deleting the variable. If the prediction error is not increasing after deleting the variable, it can be deleted from the network input; otherwise, it will be retained in the network input. With this method, improved accuracy can be achieved by deleting the irrelevant variables and keeping relevant variables in the network input.

3.2 Robust Regression Based on M-estimation

Suppose that we have obtained the nonlinear principal component scores $T \in R^{N \times k}$ and the variable $y \in R^{N \times 1}$ to be predicted satisfies the model

$$y = T\theta + e \tag{10}$$

where θ is an unknown vector parameter and e is an independent random variable.

Let $r = y - T\hat{\theta}$ denote the residuals from the estimated $\hat{\theta}$, the objective of M-estimators is

$$\min \sum_{i=1}^{N} \rho(r_i) \tag{11}$$

The function ρ is defined as follows

$$\rho(r) = \begin{cases} \frac{1}{2} r^2 & |r| \le \gamma \\ \gamma |r| - \frac{1}{2} \gamma^2 & |r| > \gamma \end{cases} \tag{12}$$

where γ is a given positive number.

4 Case Study

The proposed method is applied to the data from an atmospheric tower which is located in the refinery plant in Dushanzi, Xinjiang. The end boiling point of the top draw is predicted by the proposed method and conventional linear PCA method and neural network method respectively.

There are 150 observations, where 120 observations are chosen as the training set and 30 as the validation set.

All variables in the training data set are mean centered and variance scaled a priori in order to give equal weights to each. Variable selection is implemented first to yield a 11 relevant variables set. In the training step, mean centered and scaled data are then exposed to the NLPCA network, which has 8-4-10-11 neurons in each layer, yielding a 4 principal component model. For comparison purposes, the linear PCA was applied to the same data. As a performance criterion, the root mean square of error (RMSE) [Eq. (13)] is evaluated to compare the prediction ability of the developed 4 PCA and 4 NLPCA models on the training and

validation data.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}\sum_{j=1}^{m}\left(x_{ij} - x_{ij}'\right)^2}{mN}} \tag{13}$$

Table 1 Comparison of RMSE with the linear and nonlinear PCA model

Model	Training data set	Validation data set
Linear PCA model	0.3823	0.3896
Nonlinear PCA model	0.3100	0.3150

The result of the comparison is given in table 1. While the linear PCA gives 0.3823 and 0.3896 RMSE on the training and validation data sets, respectively, the nonlinear PCA provides 0.3100 and 0.3150 RMSE. This suggests that in order to capture the same amount of information, the linear PCA entails utilization of more principal components than its nonlinear counterpart. As a result, the information condensed in the nonlinear principal components addresses the underlying events more efficiently than the linear ones.

The comparison of the proposed nonlinear PCA method and a 4 linear PCA method and a feed-forward neural network that contains 8 hidden nodes in the models are given in table 2. RMSE is defined as follows

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}\left(y_i - y_i'\right)^2}{N}} \tag{14}$$

Table 2 Comparison of RMSE with different methods

Method	Training data set	Validation data set
Nonlinear PCA	3.6625	3.5079
Linear PCA	3.7073	3.5841
FFNN	2.9197	5.4955

The results show that the FFNN method provides the best performance for the training data set, but the performance for the validation data set is the worst, while the proposed nonlinear PCA method provides the best performance for the prediction.

5 Conclusions

Due to the nonlinear nature of most chemical processes, a soft sensor is constructed based on the nonlinear principal component analysis. An application of a crude oil atmospheric distillation tower is presented. The results and the comparison with conventional linear PCA method and feed-forward neural network method illustrate that the proposed method gives a better performance. Since the data often contain multiple scales information, a future study will be focused on the integration with multiple scale methods.

References

[1] Park S, Han C. A Nonlinear Soft Sensor Based on Multivariate Smoothing Procedure for Quality Estimation in Distillation Columns. Comput. Chem. Eng, 24: 871~877, 2000

[2] Mejdell T, Skogestad S. Estimation of Distillation Compositions from Multiple Temperature Measurements Using Partial-Least-Squares Regression. Ind Eng Chem Res, 30(12): 2543~2555, 1991

[3] Mejdell T, Skogestad S. Composition Estimator in a Pilot-Plant Distillation Column Using Multiple Temperature. Ind Eng Chem Res, 30(12): 2555~2564, 1991

[4] Mejdell T, Skogestad S. Output Estimation Using Multiple Secondary Measurements: High-Purity Distillation. AIChE J, 39(10): 1641~1653, 1993

[5] Willis M J, Massimo C DI, Montague G A, Tham M T, Morris A J. Artificial Neural Networks in Process Engineering. IEE Proceedings-D, 138(3): 256~266, 1991

[6] Willis M J, Montague G A, Massimo C DI, Tham M T, Morris A J. Artificial Neural Networks in Process Estimation and Control. Automatica, 28(6): 1181~1187, 1992

[7] Doymaz F, Chen J, Romagnoli J A, Palazoglu A. A Robust Strategy for Real-Time Process Monitoring. J. of Process Control, 11: 343~359, 2001