Proceedings of the 12th KACC, October 1997

신경회로망에 근거한 강건한 비선형 PLS

Robust Nonlinear PLS based on Neural Networks

°유준, 홍선주, 한종훈, 장근수

포항공과대학교 화학공학과, 공정산업의 지능자동화 연구센터(Tel:279-5967; Fax:279-3499;E-mail:liu@backdu.postech.ac.kr)

Abstract In the paper, we propose a new method of extending PLS (Partial Least Squares) regression method to nonlinear framework and apply it to the estimation of product compositions in high-purity distillation column. There have been similar efforts to overcome drawbacks of PLS by using nonlinear-mapping ability of neural networks, however, they failed to show great improvement over PLS since they focused only on capturing nonlinear functional relationship between input and output data, not on nonlinear correlation in the data set. By incorporating the structure of Robust AutoAssociative Networks (RAAN) into that of previous nonlinear PLS, we can handle nonlinear correlation as well as nonlinear functional relationship. The application result shows that the proposed method performs better than previous ones even for nonlinearities caused by changing operating conditions, limited observations, and existence of measurement noises.

Keywords PLS, Neural networks, Nonlinear correlation, Composition estimator, Distillation column

1. Introduction

Recently, great advances have been made in the development of the analytical instrumentation and sensors to provide on-line measurements, but the new sensors have not been preferred yet because they still suffer from large measurement delays, narrow operating conditions, high investment/maintenance costs, low reliability and so on.

For these reasons, there have been attempts to overcome the problems presented by the lack of on-line quality measurement because infrequent product quality measurement imposes severe limitations on achieving desirable control performance. In that attempts, state estimation method and soft-sensor techniques have been proposed, aimed at inferring difficult-to-measure and unmeasurable quality variables [1]. However, since the former is highly dependent upon the availability of a representative mathematical model and the measurement of secondary variables with low noise corruption, data based models that are based on input-output data pair are frequently used as an alternative [1]. These non-mechanistic model approaches can form the basis of soft-sensors. In these non-mechanistic models, regression technique is often used to infer the relationship between input and output data. It is shown that partial least squares (PLS) provides a general method for building empirical inferential models when one has data on a large number of process variables and when these variables are highly correlated with one another. However, since PLS is linear method, there are severe limitations especially when correlation in the data set and functional relationship between input and output data are nonlinear.

Another modeling method employed in data based models is artificial neural networks. It has been shown, using the Stone-Weierstrass theorem, that a two-layer feedforward network with an arbitrary large number of nodes in the hidden layer can approximate any continuous function to a desired accuracy [2], and the direct neural networks approach performs much better than linear techniques in some cases.

However, it has similar problems to the ordinary least-squares method particularly in the case that process has limited data. The number of weights in a multilayer network of m inputs and p outputs could be larger than the number of observations. Therefore, some of the weights cannot be uniquely determined from the observed data. With problems mentioned above there have been efforts to take advantages of the two methods-PLS and neural networks [3,4], but these efforts failed to show great improvement over standard linear PLS for most cases because they ignore nonlinear correlation in the data set.

In this paper, we present a new method of integrating PLS and neural networks and apply it to estimating the product compositions in high-purity distillation columns using multiple temperature measurements. The method uses the universal approximation property of neural networks to extend the standard linear PLS modeling method to a nonlinear framework. The resulting nonlinear PLS model can capture nonlinear functional relationship between input and output data without the loss of PLS's generalization property. In addition, we incorporate the structure of Robust AutoAssociative Networks (RAAN) into this nonlinear PLS so that by use of the nonlinear PCA ability of RAAN, we can capture nonlinear correlation in the data set without any difficulties. The Robust Nonlinear PLS (RNPLS) that we propose exceeds existing neural networks combined with PCA as well as standard linear PLS in terms of nonlinear-mapping ability, noise suppression, and robustness to sensor failure and capturing nonlinear correlation.

2. Current approaches for inferential model building

In the followings, current methodologies used for obtaining inferential model based on input-output data are presented. Here we consider building only static inferential model. Static modeling is quite simple since it requires little modeling efforts compared to dynamic one, and for special cases one of which is our example discussed later, it can also model the dynamic behavior although the dataset used to obtain the model is static. Undoubtedly, the methodologies can also be used to build dynamic model as well as static one.

2.1 PLS(Partial Least Squares or Projection to Latent Structures)

Inferential model building is based on a reference (calibration) data set which can be separated into two matrices; a matrix **X** associated with the process measurements and a matrix **Y** associated with the quality measurements which are not generally available in on-line. The objective is to develop an inferential model that can predict current (or future) values of quality variables using current (or future) measurements of the process variables.

Here, we briefly explain basic idea of PLS. For detailed algorithm and other information of PLS, refer to Hoskuldsson [5].

In PLS regression, the latent variables are determined in order have to the largest covariance with the dependent variables. In this way, when a vector is not needed to describe a specific variable, the vector will not be used in the estimation of the variable. Therefore it is possible to describe a variable using few components.

The application area of PLS is quite wide - from analytical chemistry to chemical engineering. In chemical engineering, the most successful application is composition estimation for distillation column. However, since it is linear method, using PLS in nonlinear problems can sometimes be in adequate. To avoid this severe limitation, the methods combining neural networks and PLS have began to propose.

2.2 Neural networks/PLS(NNPLS)

Many authors have proposed new methods combining PLS and neural networks to handle nonlinearity as well as correlation between data since neural networks have similar problems as ordinary least-squares in the case of correlated inputs and limited data, although it shows better performance than PLS. The early ones are to combine Principal Component Analysis (PCA) with neural networks [6] and the recent ones are to integrate the PLS regression and neural networks to formulate an approach which can handle nolinearity, correlated inputs and limited observations [3,4].

For example, Holcomb and Morari [4] proposed a structure of NNPLS (Fig. 1) and its training algorithm.

The Full algorithm is :

1. Perform PCA to decide number of directions.
2. Initialize feature layer with PCA directions.
3. Perform training on output layer; make no changes to feature layer.
4. Perform training with full networks including output layer.
5. If performance unsatisfactorily, choose new feature layer initial values and go to step 3.

However, their method as well as other NNPLS failed to show great improvement in their performance over linear PLS for most cases and showed nearly same performance to each other. The reason is that they focused only on capturing nonlinear functional relationship between input and output data, not on nonlinear correlation in the data set, which is another important drawback of linear PLS.

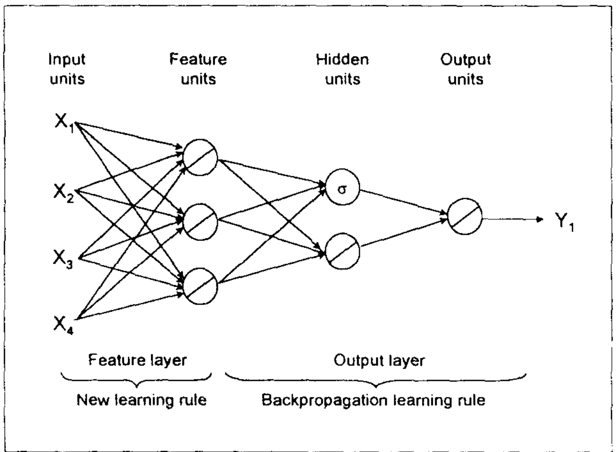

Fig. 1 Neural networks/PLS structure proposed by Holcomb and Morari [4]. σ 's and diagonals denote sigmoidal nodes and linear nodes respectively.

3. Robust Nonlinear PLS(RNPLS)

3.1 Robust Autoassociative Networks

Nonlinear principal component analysis(NLPCA) is a novel technique for multivariate data analysis, similar to PCA. NLPCA, like PCA, is used to identify and remove correlations among problem variables as an aid to dimensionality reduction, visualization, and exploratory data analysis. While PCA identifies only linear correlations, NLPCA uncovers both linear and nonlinear correlations, without restriction on the character of the nonlinearities in the data. Kramer [7] shows that five layer feedforward networks(not three layer networks) is adequate for NLPCA and proposes guideline of determining number of hidden nodes. In addition, when it is trained properly, it can be used to preprocess data so that sensor-based calculations can be performed correctly even in the presence of large sensor noises, biases, and failures. Due to this property, Kramer names it Robust AutoAssociative networks (RAAN). Its detailed structure is given in Fig. 2.

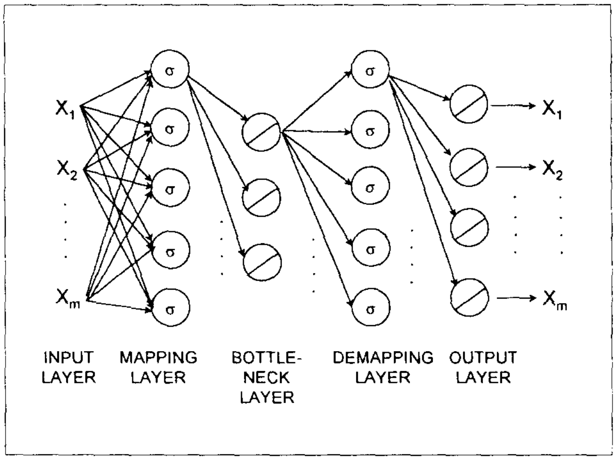

Fig. 2. Networks architecture for NLPCA using autoassociative networks. σ 's and diagonals have same meaning as in Fig. 1.

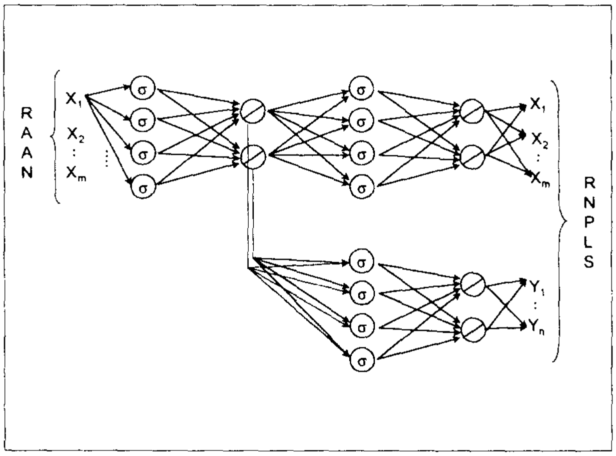

Fig. 3. Proposed RNPLS's architecture

3.2 RNPLS

As mentioned above, using linear PCA in nonlinear problems can sometimes be inadequate. For example, it has been shown that if PCA is used in nonlinear problems, minor components do not always consist of noise or unimportant variance, but they contain important information[8]. If the minor components are kept, the PCA might contain too many components to be useful for solving an application problem. NLPCA can handle this problem with fewer components because it estimates a curve or surface passing through the middle of the observations. Therefore incorporating structure of RAAN into that of previous NNPLS, it can capture nonlinear correlations as well as show better performance with fewer components than previous one. Resulting structure of RNPLS is given in Fig. 3.

4. Examples - Composition estimator

As mentioned above, same situation occurs in distillation column control. Therefore the most popular alternative to analyzers is to use secondary measurements, which is able to infer product composition. Since tray temperature measurement is reliable, inexpensive and has negligible measurement delays, it has been frequently used. Although, for the binary column at constant pressure, the temperature at the column end is an exact indicator of the product composition, the use of a single temperature at the column end section is generally not reliable for many reasons. In late 80's, multivariable regression techniques (PCR and PLS) began to get an attention within community and it was found that by using all available temperature measurements, PLS-based estimator shows good performance compared to other estimators such as the dynamic Kalman filter and the static Browsilow inferential estimator [1]. It also enhances the performance through the effective transformation and scaling of the process variables but makes the estimators sensitive to noise [1,9].

In following sections, we will present the definition of composition estimator, evaluation criteria developed to evaluate estimator's performance, and variable transformation techniques used to reduce nonlinearity.

4.1 Problem Definition.

Consider the binary distillation column with constant pressure, and feed and reflux stream as saturated liquid. Then specifying each value of feed composition z_F, distillate composition y_D, and bottom product composition x_B yields unique steady-state profile of the tray temperatures. The objective is to obtain the best estimate $\hat{\mathbf{y}}$ of the product compositions using these steady-state tray temperatures, θ. The general form of the estimator may be written as

$$\hat{\mathbf{y}} = \mathbf{K}(\theta)$$

where $\hat{\mathbf{y}} = \begin{pmatrix}\hat{y}_D & \hat{x}_B\end{pmatrix}^T$ and the $\mathbf{K}(\cdot)$ becomes constant matrix for PLS and nonlinear function matrix for NNPLS and RNPLS. For binary distillation column with n-trays, the dimension of matrix $\mathbf{K}$ is $2 \times (n+1)$ and the problem is to find optimal values of $2(n+1)$ parameters.

4.2 Evaluation criteria

Prediction Error Sum of Squares (PRESS) (Montgomery, 1992) is used to evaluate the absolute performance :

$$\text{PRESS} = \sum_{i=1}^{N} (\hat{y}_i - y_i)^2$$

4.3 Variable transformation techniques

Since the composition and temperature profiles are nonlinear functions of the operating variables, many attempts have been made to overcome this nonlinearity. A simple and efficient way is to use nonlinear transformations on each variable. Logarithmic transformation of the product compositions has been proposed by several authors [9,10] as an effective way to linearize the static as well as dynamic response. For binary mixtures the following transformation is used :

$$Y_D = \ln\left(\frac{y_D}{1 - y_D}\right)$$

where y_D is the distillate composition. Various transformation techniques were investigated by Mejdell and Skogestad [9] and it was proposed that logarithmic transformation of both the composition and the temperatures improve the estimates by linearizing these responses and profiles. The proposed transformation is

$$L = \ln\left(\frac{\theta - \theta_L}{\theta_H - \theta}\right)$$

where θ is the tray temperature, and θ_L and θ_H are some reference temperature. One may use the temperatures at the column ends when there are two additional measured reference temperatures, or θ_L and θ_H are the boiling temperatures of the pure light and heavy components, T_L^b and T_H^b respectively.

4.4 Example column

The column has 15 theoretical stages with total condenser and reboiler. The feed stream enters the column at stage 8 as saturated liquid. Binary mixture case consisting of one light and heavy key component is considered. The steady state and dynamic simulations of the column are performed using simulator, HYSYS™. The steady state and dynamic simulation conditions for the case are given in Table 1. and 2. respectively.

Table 1. Steady-state simulation conditions

	Base case condition	Variations in steady-state reference set
Inputs		
Feed flowrate	36 m^3/hr	Constant
Feed temperature	73 ℃	71 ~ 75.5 ℃
Feed composition		
Methanol	50 %	40 ~ 60 %
Water	50 %	40 ~ 60 %
Outputs		
Distillate		
Methanol	99 %	97 ~ 99.667 %
Water	1 %	0.333 ~ 3 %
Bottom		
Methanol	1 %	0.333 ~ 3 %
Water	99 %	97 ~ 99.667 %

Table 2. Dynamic simulation conditions

Tray size	
Diameter	10.1 cm
Weir height	1 cm
Condenser vessel volume	10.1 L
Reboiler vessel volume	8 L
Tower volume	20 L
Cooling volume	4.6 L
Liquid holdup time	5 min
Setpoint change	
Top	97 % → 99 % → 99.6 %
Bottom	3 % → 1 % → 0.3 %

4.5 Simulation and result

Using HYSYS™ and Table 1 as a simulation condition, we obtain 32 reference data and build 4 different models - PLS without transformation(PLS), PLS with transformation(PLS W/ TRNS), NNPLS, and RNPLS. In transformation for PLS W/ TRNS, we use logarithmic transformation of both the composition and the temperatures as Mejdell and Skogestad [9] proposed. And NNPLS model is obtained using Holcomb and Morari's method[4]. To test dynamic performance of models, we obtain dynamic data from Table 2. For both cases, we prepare noisy data which contain three different noise level (±0.1℃, ±0.2℃, ±0.3℃) to show robustness of models. to produce these noisy data, we add normally distributed noise with magnitude 0.1, 0.2, and 0.3 to reference data. Also to test estimators' own robustness to noise, we don't train NNPLS and RNPLS to these noisy data..

As a model building result, PLS and PLS W/ TRNS have 4 and 3 components respectively and networks structure for NNPLS and RNPLS are (16-3-5-2) and (16-25-2-5-2).

4.5.1 Static case

Static performances of models are given in Fig. 1 and 2. From the figures, when there is no noise in data PLS shows the worst performance and NNPLS and RNPLS show the best performance and RNPLS has fewer components than NNPLS by 1. However, as the noise level increases, performances of PLS W/ TRNS and NNPLS become worse whereas RNPLS does not. This is because RAAN has noise suppression ability at the bottleneck layer [7]. From the view of prediction power and robustness, RNPLS is the best for static case.

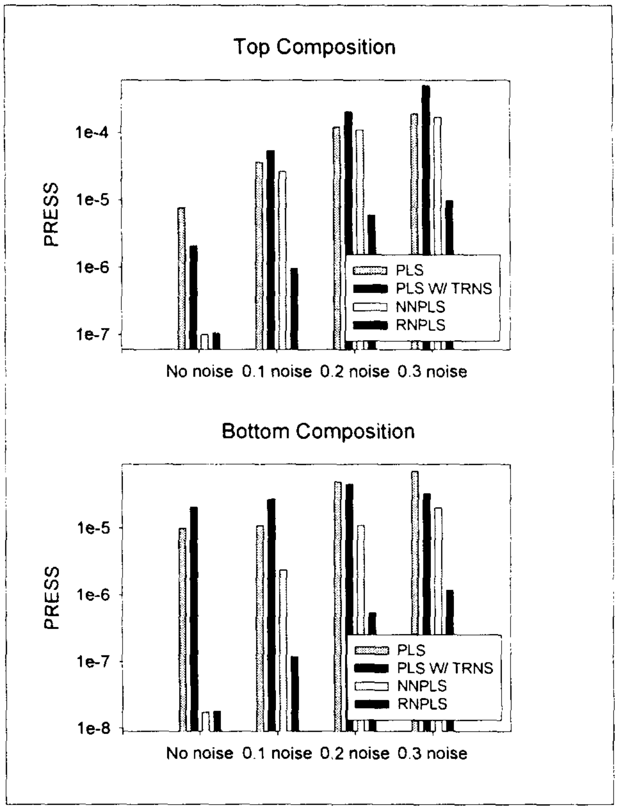

Fig. 4. Static performance of estimators to steady state reference data

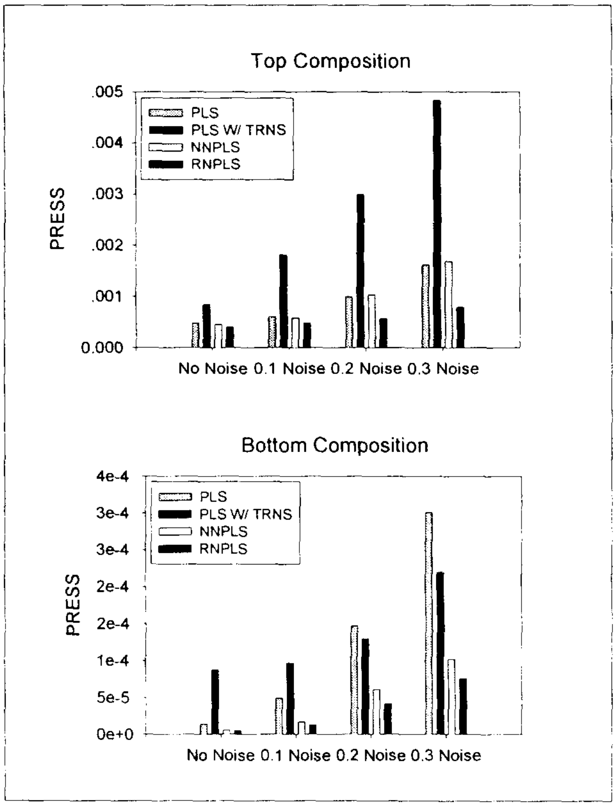

Fig. 5. Dynamic performance of estimators to dynamic data

4.5.2 Dynamic case

For most cases, dynamic performance of estimator is more important than the static because inferential model is frequently used for control purpose. Therefore, in addition to good static performance, good dynamic performance is inevitable. The result of dynamic performance test using dynamic data (setpoint change data) shows that prediction power and robustness of RNPLS exceeds all others. Especially for top composition estimation which is highly correlated, RNPLS performs better than PLS W/ TRNS and NNPLS. This can be explained that RAAN captures non-linear correlation, which cannot be captured by others since it is minor but has important information about system's dynamics.

5. Conclusion

In this work, a new method for extending linear PLS to nonlinear framework is proposed. Proposed method (RNPLS) can capture nonlinear correlations with fewer components in addition to nonlinear mapping and application result show that its prediction power surpasses others even for dynamic estimation and existence of measurements noise. Especially when it is used for control, only RNPLS can guarantee good control performance since it shows excellent performance for dynamic case compared t others.

6. Acknowledgement

Financial support from the Korea Science and Engineering Foundation (KOSEF) through the Automation Research Center (ARC) at Pohang University of Science and Technology (POSTECH) is gratefully acknowledged.

7. Reference

1. T. Mejdell and S. Skogestad, "Output Estimation Using Multiple Secondary Measurements: High-Purity Distillation", *AICHE J.*, Vol. 39, No. 10, pp. 1641-1653, 1993.
2. K. Hornik, M. Stinchcombe and H. White, "Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks", *Neural Networks*, Vol. 3, pp. 551-560, 1990.
3. S. J. Qin and T. J. McAvoy, "Nonlinear PLS Modeling Using Neural Networks", *Computers chem. Engng*, Vol. 16, No. 4, pp. 379-391, 1992.
4. T. R. Holcomb and M. Morari, "PLS/Neural Networks", *Computers chem. Engng*. Vol. 16, No. 4, pp. 393-411, 1992.
5. A. Hoskuldsson, "PLS regression methods", *J. Chemometrics*, Vol. 2, pp. 211-228, 1988.
6. S. J. Orfandis, "Gram-Schimidt Neural Networks", *Neural Computation*, Vol. 2, pp. 116-126, 1990.
7. M. A. Kramer, "Autoassociative neural networks", *Computers chem Engng*. Vol. 16, No. 4, pp. 313-328, 1992.
8. L. Xu, E. Oja, C. Y. Suen, "Modified hebbian learning for curve and surface fitting", *Neural Networks*, Vol. 5, pp. 441-457, 1992.
9. T. Mejdell and S. Skogestad, "Estimation of Distillation Compositions from Multiple Temperature Measurement Using Partial-Least-Squares Regression ", *Ind. Eng. Chem. Res.*, Vol. 33, pp. 2543-2555, 1991.
10. B. Joseph, C. B. Browsilow, "Inferential Control of Processes", *AICHE J.* Vol. 24, pp. 485-509, 1978.