

## 내적 조합 모델 PLS를 이용한 소프트 센서 설계

홍 선 주, 한 증 훈

포항공대 화학공학과, 공정 산업의 지능 자동화 연구센터  
(1998년 11월 6일 접수, 1998년 11월 17일 채택)

### Soft sensor design based on PLS with hybrid inner model

Sun Ju Hong, Chong Hun Han

Department of Chemical Engineering, Automation Research Center,  
Pohang University of Science and Technology

(Received 6 November 1998; Accepted 17 November 1998)

#### 요 약

온라인 가스 크로마토그래피와 같은 실시간 분석기기는 데이터 채취 및 분석과정의 시간이 오래 걸려 측정 신뢰성이 낮으며 또한 설치 및 보수비용이 많이 들어서, 실제 가스 산업 공정에서는 온도, 압력과 같은 공정의 운전 변수들을 이용하여 간접적으로 가스의 성분 및 양을 감지하는 소프트 센서에 관한 연구가 활발하게 진행 중이다. 본 연구에서는 데이터 기반 소프트 센서의 설계에 있어서 공정 데이터가 갖는 패턴을 잘 반영하여 추정 능력을 높이는 내적 조합 모델 PLS를 제안하고, 이를 산업체의 농도 추정 목적으로 적용하였다. 제안된 PLS모델은 기존에 제시된 모델들보다 뛰어난 추정 능력을 보였다.

**Abstract** - It takes quite a long time for an analyzer, such as gas chromatography, to measure a bulk property of a system, which prevents on-line measurements. Also, the cost of installation and maintenance is very high. Consequently, some other means is needed for on-line measurements of properties and the development of soft sensors based on process variables like temperature and pressure is of great interest. In the field of gas industry, the development of a soft sensor which makes indirect on-line measurements of gas compositions and flow rate, is in progress. In this paper, we proposed a hybrid inner model PLS which improved the prediction performance by taking into account the data structure, as an empirical modeling algorithm. When applied to a design of a soft sensor of a distillation tower, the hybrid inner model PLS showed better prediction performance than other methods.

**Key words** : Process variable, On-line measurement, Data structure, Soft sensor, Hybrid inner model PLS

#### 1. 서 론

최근에는 지구환경을 보호하기 위해 ISO14000에 대응하는 것은 기업의 사회적 책임을 수행하는 의미에서도 중요한 목표로 되고 있다. 따라서 각 공장에서도 환경에 미치는 영향을 관리하고 또한 건전한 환경을 계속적으로 유지하기 위해 감시용 분석계가 많이 사용되고 있

다. 예를 들어, 화학플랜트에서 부산물로 방출될 수 있는 질소산화물(Nox), 황산화물(Sox) 등의 독소가스 누출의 조기감지 및 정확한 양의 추정은 환경측면뿐만 아니라, 안정된 조업을 수행하는 데 있어서도 매우 중요하며, 배수처리 플랜트에서 바다나 하수 등으로 방출하는 처리수의 배수기준의

엄격한 관리는 깨끗하고 안전한 수질공급 측면에서 중요시되고 있다. 하지만, 공정에서 부산물로서 배출되는 인체에 유해한 가스나, 회분식 폐수처리시스템의 폐수의 법적 기준 값의 예측은 GC(Gas Chromatograph)등의 실시간 분석기의 낮은 신뢰성 및 비싼 가격 등으로 인하여 직접적인 실시간 추정이 어려운 실정이다.

따라서, 독소가스, 폐수와 폐기물 등의 좀더 신속하고 신뢰성 있는 정성적, 정량적 예측을 위하여 최근에 많은 환경플랜트에 설치되고 있는 DCS(Distributed Control System)와 PIS(Plant Information System)를 이용하여 공정에 대한 막대한 양의 데이터를 실시간으로 받아들이고 또한 방대한 과거의 데이터(historic data)를 저장함으로써, 플랜트의 과거와 현재에 걸친 많은 양의 상관된 데이터를 이용하여 추정하고자 하는 변수들, 즉 가스, 폐수 또는 폐기물의 정성, 정량 값을 간접적으로 추정하는 방법들이 모색되고 있다.

이처럼, 공정데이터를 이용하여 실시간으로 신속하고 정확하게 추정하기 어렵거나 불가능한 품질변수(가스, 폐수 또는 폐기물 등의 정성, 정량 값)를 예측하는 모델을 소프트센서라고 하며, 이는 공정 데이터를 이용하여 품질변수를 추정하는 감시(monitoring)를 통해 이상을 조기에 발견(fault detection), 진단(diagnosis)하는데 응용할 수 있다. 한편, 소프트센서를 이용한 추론제어(inferential control)는 이미 증류공정이나 생물환경공정 및 고분자공정에 널리 응용되고 있다[4],[5]. Fig 1은 소프트 센서를 이용한 추론제어의 schematic이다

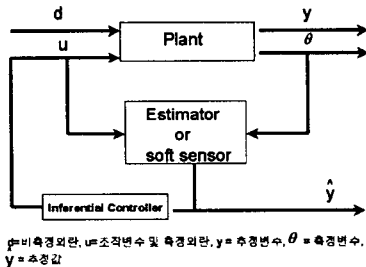


Fig 1. Soft sensor schematic for inferential control

근래 들어 데이터 기반 소프트 센서 기술로써 Principal Component Regression(PCR), Partial

Least Square or Projection to Latent Structure(PLS) 등 다변량 통계적 분석 방법이 많이 도입되고 있는데, 이들은 서로 강하게 상관되어 있는 공정 데이터를 효과적으로 다루면서 공정변수와 추정변수와의 관계를 맺는 방법이다. 이 방법은 방대한 양의 데이터를 간단한 통계적인 방법을 이용하여 쉽게 모델링할 수 있고, 동시에 공정의 데이터 분석에 이용할 수 있는 도구를 얻을 수 있다는 점에서 매우 유용하다[3].

이들은 서로 상관되어 있는 데이터를 데이터 공간에 대해 분포가 넓은 축부터 시작하여 순차적으로 서로 직교하는 주성분(Principal Component)을 구하여 그 중에 의미 있는  $a$ 개의 주성분으로써 시스템을 근사하는 방법이다. 이 중 PLS방법은 주성분을 구할 때, 입출력 score 벡터의 내적 모델링을 통해 입력 데이터 행렬이 출력 데이터 행렬과 가장 큰 covariance를 갖도록 반복적으로 주성분을 재배열하는 과정을 갖도록 하는 것이 그 특징이다[1].

PLS는 공정 변수들간의 redundancy를 해결하고 측정 잡음(measurement noise)을 없애 주면서, 보다 나은 입출력 관계를 얻을 수 있어서 근래 들어 많은 이론적 연구와 함께 실제 공정의 적용에 중요한 현안이 되고 있다.

그러나 PLS 알고리즘은 선형 방법으로써 강한 비선형성, 복잡성과 불확실성을 가지고 있는 공정의 추정 모델로는 부적합해서, 근래 들어 입출력 score 벡터를 polynomial, 신경회로망 등의 비선형 함수로써 회귀의 성능을 높여 모델의 performance를 높이려는 시도가 활발하게 연구되고 있다[2]. 이 때 내적 모델링 방법으로써 비선형 함수를 사용하고 있어서 이렇게 개발된 모델들을 비선형 PLS라고 불리 운다[6],[7].

선형 PLS와 비선형 PLS는 입출력 데이터의 주성분을 정의하는 단계마다 일률적으로 선형 함수 또는 비선형 함수를 사용하고 있기 때문에 주성분을 정의하는 모든 단계마다 최적으로 입출력 score 벡터가 모델링 되고 있다고 볼 수 없다. 즉, 비선형성을 띄는 공정에 선형 PLS를 이용하여 모델링을 할 경우 정보의 손실(loss)이 필연적이며, 비선형성이 약한 공정에 비선형 PLS를 도입하게 되면 측정 오차나 특이점(outlier) 등에 민감하여 모델의 신뢰성이 떨어진 다.

따라서, 본 연구에서는 데이터 기반 소프트 센서의 설계 기술로써 입출력 데이터의 구조를 고려하여 최소 자승법 및 신경 회로망을 선별한 최적 내적 조합 모델 PLS 소프트 센서를 제안 하였고, C6/C7 splitter column의 담자 생산품 인 톨루엔의 조성 추정 소프트센서 설계 기술에

적용하였다.

## 2. 내적 조합 모델 PLS 소프트 센서 설계 과정

내적 조합 모델 PLS 소프트 센서는 크게 data step과 model step의 두 가지 과정을 거쳐서 설계 된다.

### 2-1. Data step

Data step에서는 공정의 정상 상태 데이터 수집과 수집된 데이터 중에 추정하고자 하는 변수와 상관관계가 있는 공정 변수를 선정하게 되는데, 변수를 선정하기에 앞서 변수들은 자동 크기조절(auto scaling)을 거친다.

본 연구에서는 수집된 데이터 중 정상 조업 영역에서 많이 벗어나는 관측치는 PCA를 통해 제거하였고, 그 외에 측정 오차나 센서의 고장으로 인한 잘못된 분석 데이터는 모델에 참여시키지 않았다.

추정하고자 하는 변수와 연관성을 갖는 공정 변수의 선정은 PLS의 회귀 계수로써 0근방의 값을 갖는 변수들은 PLS모델에 기여하는 정도가 거의 없다고 판단하여 모델에서 제외시켰다.

### 2-2. Model step

Model step은 대상 공정에 적합한 추정 모델을 세우고, 그 모델의 검증이 수반되는 단계이다. 일반적으로 PLS모델은 시스템을 가장 비슷하게 근사하는 최적의 주성분의 개수를 결정함으로써 검증된다.

본 연구에서 제안하는 내적 조합 모델 PLS는 선형 함수인 최소 자승법과 비선형 함수인 신경 회로망을 PLS의 내적 모델로써 적절하게 조합하는 알고리즘으로써 내적 모델의 구조와 주성분의 개수는 cross validation을 통해 결정되어진다.

모델의 goodness of fit은 FVU(Fraction of Variance Unexplained)로써 평가하였고, 모델의 예측성능은 RMSEP(Root Mean Squares Error of Prediction)를 이용하여 선형 PLS, 비선형 PLS 및 신경 회로망과 비교하였다.

$$FUV = \frac{SSE}{SST} = \frac{\sum_{i=1}^n \sum_{j=1}^m (y_{ij} - \hat{y}_{ij})^2}{\sum_{i=1}^n \sum_{j=1}^m (y_{ij} - \bar{y}_i)^2}$$

$$RMSEP(i) = \sqrt{\frac{1}{n} \sum_{j=1}^n (j_{ij} - \hat{y}_{ij})^2}$$

## 3. 내적 조합 모델 PLS

### PLS

$$X = t_1 p_1^T + t_2 p_2^T + \dots + t_a p_a^T = TP^T + E \quad (1)$$

$$Y = u_1 c_1^T + u_2 c_2^T + \dots + u_a c_a^T = UC^T + F \quad (2)$$

$$U = b_1 t_1 + b_2 t_2 + \dots + b_a t_a = BT + R \quad (3)$$

X는 공정 변수인 온도, 압력 등과 같은 측정 변수를 가리키며, Y는 추정하고자 하는 가스의 조성, 유량 등의 추정 변수이다.

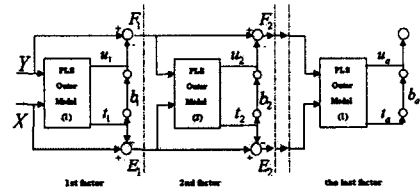


Fig 2. The schematic of PLS algorithm.

PLS 모델은 X 블록의 scores와 Y 블록의 scores가 가장 간단한 형태의 회귀로 이루어져 있으며, 각각 블록 안에서의 외적변환(outer relation)과 두 블록의 관계를 위한 내적변환의 두개의 구조를 이룬다(Fig 2).

식 1), 2)는 각각 X, Y블록에 대한 외적변환을 나타내며, 식 3)은 입력력 score t와 u가 최소 자승법으로써 관계를 맺는 내적 변환을 의미한다. 그리고 여기서 E, F, R은 잔차 행렬(residual matrix)들이고 E와F가 최소가 되어 random error 라고 판단되어 질 때까지, 위와 같은 과정이 반복 계산된다[1].

이 때 비선형 PLS는 내적 변환을 신경회로망 등의 비선형 함수로써 비선형 근사를 하는 것을 말하는데, 식 4)는 신경 회로망을 이용한 내적 모델링을 나타내고 있다

$$u_h = N(t_h) + r_h \quad (4)$$

N()은 신경 회로망에 의해 표현된 비선형 관계를 나타내고 있다.

이 방법은 PLS 외적 변환에 의해 데이터가 전처리가 된 후에 network에 의한 학습이 이루어

어지므로 신경 회로망 PLS의 network은 SISO(single input single output)가 되고, 이것은 가중치(weights)수를 현저하게 줄임으로써 over-parameterization을 피할 수 있다. 또한 국부 최소값의 수도 network의 크기가 작음으로 인해 더 줄어들기 때문에 일반적인 신경 회로망 모델링 방법보다 더 강건하고 예측 성능이 나은 모델을 설계할 수 있다[2].

PLS는 정의된 주성분들 가운데 의미 있는 a 개로만 시스템을 근사하는 방법이므로, 어느 정도인 공정의 정보 손실을 가져오게 된다. 또한 기존의 PLS는 주성분을 정의하는 각 단계의 입출력 score벡터의 관계와는 무관하게 일괄적으로 내적 모델 함수가 선형 함수 또는 비선형 함수로써 정해져 있으므로 각 단계마다 다른 크기의 모델 오차가 필연적으로 생긴다.

내적 조합 모델 PLS

본 연구에서 소프트 센서 기술로써 제안하는 내적 조합 모델 PLS는 각 단계마다 주성분을 정의할 때의 내적 모델의 오차를 최소화하여 공정의 정보의 손실을 최소화하고자 각 단계마다 입출력 score 벡터의 관계를 cross validation을 통해 PRESS가 적은 함수형태를 선택하는 방법이다.

즉, h번째 단계에서의 입출력 score벡터는 일단 n-1개의 모델 집합과 한 개의 시험 데이터(test data)로 나누어진 후 식 3)과 4)의 방법으로 모델이 구성되고, 모델에 참여하지 않은 한 개의 시험 데이터는 구해진 모델을 이용하여 실제 값을 예측하고, 그에 대한 PRESS를 구한다. 함수의 결정은 이와 같은 과정을 n번 반복함으로써 식 3)과 4)에 따르는 cumulative PRESS가 작은 모델 형태를 선택한다.

**4. 사례 연구(case study)**

C6/C7 splitter process 개요

Splitter는 60단을 가지고 C6와 C7을 분리해 내는 공정이다. 현재 탑저의 톨루엔 조성을 제어하기 위해 42단의 온도를 제어변수로 사용하고 있으며 환류량과 재비기의 열용량이 조작 변수로 사용되고 있다.

측정되는 변수로는 탑상 압력, 탑저 압력, 탑상, 13단, 42단, 56단, 탑저의 온도, 그리고 원료(Feed), Distillate, Bottom flow rate 등의 13개이다.

본 사례 연구의 목적은 42단의 온도를 포함한 측정 변수들을 이용하여 탑저의 톨루엔 조성을 추정하는 소프트 센서를 설계하는 것이다.

측정 변수들은 30분 동안의 측정 치들을 평균한 값이며, 톨루엔의 조성은 하루에 두 번씩 8시간과 16시간의 간격을 두고 가스 크로마토그래피로 분석이 되고 있다. 2개월 동안의 조업 데이터로서 총 138개의 데이터 중 비정상 조업 데이터를 제거한 66개의 관측수로 모델을 구성하였다. 그 중에서 56개는 모델을 만드는데 사용하였으며 나머지 10개는 모델의 예측 성능을 위해 사용되었다.

모델링 결과 및 고찰

구해진 모델이 실제 값을 설명하지 못하는 variance 정도를 FVU를 이용하여 평가하였고, 모델 검증에 위하여 시험 데이터와 실제 값의 RMSEP으로써 각 모델의 예측 성능을 비교하였다(Fig 3, Fig 4).

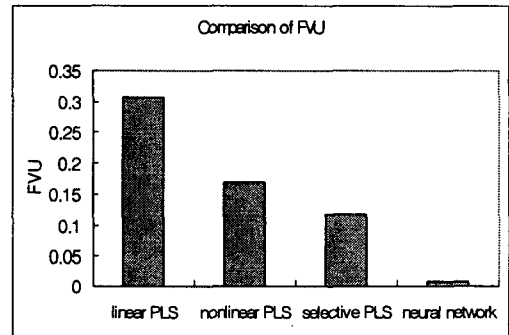


Fig 3. Comparison of FVU

FVU를 비교해 볼 때 신경 회로망이 실제 값을 가장 많이 설명하고 있음을 볼 수 있다

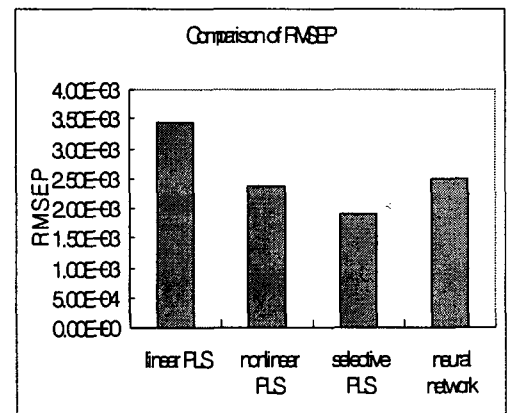


Fig 4. Comparison of RMSEP

(Fig 3). 그러나, 모델의 예측 성능면에서는

RMSEP가 가장 적은 모델링 방법은 선형 회귀와 비선형 회귀가 적절하게 조합된 내적 조합 모델 PLS 이다(Fig 4) .

이것은 증류 공정의 비선형 동적 거동성으로 인하여 신경회로망이 모델의 적합도(goodness of fit)는 가장 뛰어나지만, 모델 데이터의 수가 충분히 크지 않고 정상조업 데이터의 가능한 variation이 모두 포함되어 있지 않음으로 인해서 시험 데이터의 예측 추정에서 overfitting이 발생한 것으로 보인다. 또한 비선형 PLS도 신경회로망과 비슷한 문제점을 가지게 되면서 각 주성분의 내적 모델 단계에서 과도한 학습으로 인한 overfitting으로 시험 데이터의 예측 결과가 좋지 못하다. 한편, 외적변환을 거친 입출력 score 벡터의 적절한 선형/비선형 회귀모델링 방법은 모델링 면에서는 신경 회로망과 거의 차이가 없지만, 시험 데이터의 예측 결과는 가장 정확하게 톨루엔의 농도를 추정 하고 있음을 알 수 있다.

## 5. 결 론

가스 크로마토그래피 등과 같은 실시간 분석기의 여러 가지 문제점으로 인하여, 실시간으로 측정하기 어렵거나 불가능한 품질변수의 측정을 공정 변수를 이용하여 간접적으로 추정하는 소프트 센서는 최근 공장의 환경규제가 엄격화 되면서 공정의 부산물로 생산될 수 있는 질소산화물이나 황산화물 등과 같은 독소가스의 조기 감지에 이용되고 있다. 또한 그 외에도 증류 공정, 고분자 공정, 생물공정 등의 다양한 산업체 공정에 이용되고 있다.

본 연구에서는 선별된 내적 모델을 이용한 내적 조합 모델 PLS 소프트 센서를 제안하였는데, 이는 많은 양의 상관된 데이터를 효과적으로 모델링하는 강건한 선형 PLS와 신경회로망을 이용하여 비선형성을 적절하게 다루는 비선형 PLS를 조합한 알고리즘이다.

이렇게 설계된 소프트 센서는 C6/C7 splitter column의 탐저 생산품인 톨루엔 농도 추정을 위한 소프트 센서 설계에 적용되어 선형PLS, 비선형PLS과 신경 회로망보다 좀더 강건하고 정확하게 시험 데이터를 추정하고 있음을 입증해 보였다.

## 6. 감사의 글

본 연구를 위해 포항공대 공정산업의 지능자동화센터를 통해 재정적 지원을 해 주신 한국과학기술단에 감사를 드립니다.

## 사용 기호

$i$  :  $i$ 번째 관측 값

$n$  : 전체 관측 수

$j$  :  $j$ 번째  $y$ 변수

$p$  : 전체  $y$ 의 개수

$a$  ; 주성분 개수 중에 시스템 근사에 사용되는 개수

$h$  :  $h$ 번째 주성분 개수

## 7. 참고문헌

1. J. V. Kresta, T. E. Marlin and J. F. MacGregor, Development of Inferential Process Models Using PLS, Computers chem. Engng., vol.18, no. 7, pp. 597-611(1994).
2. S. J. Qin and T. J. McAvoy, Nonlinear PLS Modeling Using Neural Networks, Computers chem. Engng., vol.16, no.4, pp. 379-391(1992).
3. Neter, J., Wasserman, W. and Kutner, M.H. Applied Linear Statistical models, 4thed(1996).
4. Ming T. Tham, Gary A. Montague, A. Julian Morris, and Paul A. Lant, Soft-sensors for process estimation and inferential control, J.Proc.Cont, Vol 1, pp. 3-14(1991).
5. Gang Chen and Thomas J. McAvoy, Process Control utilizing Data Based multivariate Statistical Models, The Canadian Journal of Chemical Engineering, Vol 74, pp. 1010-1024(1996).
6. S. J. Qin and T. J. McAvoy, Nonlinear FIR Modeling via a Neural Net PLS Approach, Computers chem. Engng., vol.20, no.2, pp. 147-159(1996).
7. E.C.Malthouse, A.C.Tamhane and R.S.H.Mah, Nonlinear partial least squares, Computers chem. Engng, no.8, pp875-890(1997)