

RBF와 온라인 SOFM에 기반한 비선형 부분최소자승법을 이용한 가스공정의 이상감지

유태호, 손정현*, 송상옥, 윤인섭

서울대학교 응용화학부

(주)AID*

Gas Process Fault Detection Using Non-Linear Partial Least Square Method Based on RBF and On-Line SOFM

Tae Ho Lyoo, Jung Hyun Sohn*, Sang Ok Song, En Sup Yoon

School of Chemical Engineering, Seoul National University

AID Corp.*

1. 서론

다른 산업 시설물과 비교하여 보았을 때 가스공정이 가지는 특징으로는 반응 (reaction)과 비선형성(nonlinearity)을 들 수 있다. 가스공정에서의 이상발생 원인으로서는 센서와 같은 하드웨어 장치들의 물리적인 고장, 공정의 외부에서 발생하는 비정상적인 상황에 의해 공정으로 도입되는 외란 및 조업자의 기술 부족으로 인한 부적절한 조작과 같은 인적인 과실 등으로 대별할 수 있다.

비선형성을 가지는 가스공정을 제어하고 이상을 감지한다는 것은 쉽지 않은 일이나 보다 안전하고 경제적인 공정을 개발해야만 하기에 많은 이상감지 기법이 연구되고 있다.

우선적으로 가스공정 상태를 나타내는 공정 데이터를 효과적으로 처리할 수 있어야 효율적인 이상감지 기법을 개발할 수 있는 발판이 마련될 수 있을 것이다. 최근 컴퓨터의 비약적인 발전으로 여러 data mining 기법들이 효율적으로 수행되고 있다. 수많은 변수가 존재하는 가스공정은 개별적 변수 감지를 수행하는 단변량 통계분석방법을 사용하기보다는 다변량 통계분석법을 사용하여야 한다.

다변량 통계적 분석기법들은 가스공정뿐만 아니라 제약, 의약, 고분자, 생명과학과 같은 공학분야에서 널리 사용되고 있으며, 가스공정분야에서는 공정의 이상감지 등에 응용되고 있다. 그러나 이를 사용함에 있어서도 변수들간의 높은 상관성에서 기인하는 multi-collinearity 문제, 데이터 측정회수 보다 데이터를 이루는 변수가 많은 상황에서 기인하는 dimensionality 문제, 변수들간의 비선형에서 기인하는 non-linearity 문제 등의 난점이 존재한다. 이러한 난점을 극복하고자 보다 효율적인 data mining 기법을 이용하여 데이터들간의 의존성(dependency)이나 상관구조(correlation structure)와 같은 유용한 정보를 찾아 내어야 한다. 그리고 나아가 공정에 대한 해석을 쉽게 하고 획득한 데이터로부터 미지의 값을 추정함에 있어 보다 향상된 예측성능을 가지게 하여야 한다.

본 연구에서는 위에서 제시된 여러 난점들을 극복하고 보다 빠르고 정확한 가스공정 이상진단을 위해 RBF(radial basis function)와 on-line SOFM(self-organizing feature map)을 적용하여 다변량 통계기법의 하나인 부분최소자승법(non-linear partial least square)을 이용해보았다.

이와 같이 구성된 부분최소자승법 모델을 가지고 BTX공정에서 중요한 역할을 하는 스팀 생산 보일러 공정에 적용하여 보았다.

2. 본론

NLPLS(non-linear partial least square)모델

PLS(partial least squares)는 데이터의 노이즈가 심하고 변수간의 상관관계가 강하며, 제한된 수의 데이터만이 존재하는 문제들에 대해 강력한 선형회귀성능을 보이는 다변량 통계분석법의 하나이다.

이는 X블록에 대해 PCA를 수행한 후 얻어지는 PC score 행렬에 Y블록을 다중선형회귀 시키는 PCR(principal component regression)방법을 확장시켜 Y블록에 대해서도 PCA를 수행하여 score 벡터를 구하고 이것을 X블록의 score벡터와 회귀시키는 방법이다[1].

PLS는 다음과 같이 외적관계(outer relation)와 내적관계(inner relation)를 독립적으로 수행하는 회귀모델로 볼 수 있다.

- 외적관계 : X와 Y블록에 대해 PCA를 수행하여 각 score 행렬인 T, U를 얻는다.
- 내적관계 : T, U 간에 회귀모델인 MLR을 수행한다.

그러나 이러한 PLS모델은 X, Y 블록간에 비선형성이 있을 경우는 적절치 못하여서 외적관계에 비선형 PCA를 사용하고 내적관계에는 신경망모델을 이용한 NLPLS(non-linear partial least square)모델을 사용한다. NLPLS방법은 크게 내적관계에 비선형 회귀모델을 이용하는 경우와 외적관계와 내적관계에 모두 비선형 모델을 사용하는 경우로 나누어 볼 수 있다.

본 연구에서는 앞서 말한 바와 같이 선형 PLS를 확장하여 외적변환과 내적회귀모델로 비선형모델을 적용하는 비선형 PLS를 제안하였다. 즉 외적변수변환은 사영기반(projection-based)인 principal curve 알고리즘을 신경망으로 구현한 on-line SOFM(self-organizing feature map)을 이용하였고, 내적회귀모델로는 RBF(radial basis function)을 사용하였다.

즉 요약하자면 on-line SOFM을 이용한 외적변환을 통해 비선형 관계성을 가지는 입력, 출력데이터는 잠재변수로 사영되며, 이렇게 얻어진 잠재변수간의 회귀는 신경망을 통해서 이루어진다.

On-Line Self-Organizing Feature Map

본 연구에서는 on-line SOFM을 외적변환에 사용하였다.

SOFM은 Marlsburg(1973)와 Grossburg(1976)에 의해 제안되고 Kohonen에 의

해 체계가 잡혔으며 principal curve의 신경망적 구현이라 할 수 있다[2,3]. SOFM에는 패러미터의 갱신이 순차적으로 이루어지는 on-line SOFM 방식과 모든 데이터에 대한 계산이 수행된 후 패러미터 갱신이 일어나는 배치방식이 있다.

일반적으로 학습데이터가 비슷한 패턴의 많은 데이터 셋으로 구성되어있는 경우에는 배치 SOFM보다 on-line SOFM이 우수한 성능을 나타내고 있다. 또한 on-line SOFM은 사용하기가 쉽고, 규모가 큰 어려운 문제일수록 우수한 해를 구해낼 수 있다[4].

SOFM의 경우에는 최소의 직교사영거리를 갖는 점으로 입력데이터들의 사영이 이루어지는 것이 아니라 유한한 b개의 점(입력데이터의 특성 변수값을 나타낸다.) 중심으로 입력데이터의 사영이 이루어진다. 이러한 중심점들은 입력데이터의 특성 변수값을 의미하고 있다.

RBF(Radial Basis Function)

과적합(overfitting)을 방지하기 위해 $M(\ll N)$ 개의 basis 함수를 사용하는 방법으로서 일반적인 식은 아래와 같다.

$$y(x) = \sum_j^M w_j \Phi_j(x) + w_0 \quad (1)$$

여기서 $\Phi_j(x)$ 는

$$\Phi_j(x) = \exp\left(-\frac{\|x - \mu_j\|^2}{2\sigma_j^2}\right) \quad (2)$$

이며, μ_j 는 Φ_j 의 중심이다.

본 연구에서 내적 회귀모델로 사용된 RBF는 근사특성(approximation property)면에서는 다층신경망과 거의 비슷한 효율을 보이지만 신경망의 학습알고리즘이 간단하고 학습속도도 매우 빠르다는 장점이 있다[5].

3. 예제

적용된 공정은 BTX공정을 구성하는 요소 중 증기를 생성하는 메인 unit이다. 메인 unit은 증기를 생성하는 부분으로서 급수부에서 공급되는 물을 가열하여 필요한 증기를 생성하며 발생한 증기는 1차, 2차 과열기를 지나면서 고온, 고압의 증기가 된다. 두 과열기 사이에는 2차 과열기에서 나오는 증기의 온도와 압력의 조절을 용이하게 하기 위한 장치(desuperheater)가 존재한다.

사용된 센서 데이터는 실제 공장의 150 ton/h 크기의 대형 보일러 공장에서 가져온 것으로 부하 변동에 따른 변화를 보인다. 주어진 보일러의 메인 드럼에 대한 공정 데이터 중 이상이 발생한 경우의 데이터는 공정 내 연료를 공급하는 밸브의 이상으로 인한 공정상태 변화가 발생하는 부분의 데이터를 이용하였다. Fig.1 에 그 데이터의 일부를 보였다.

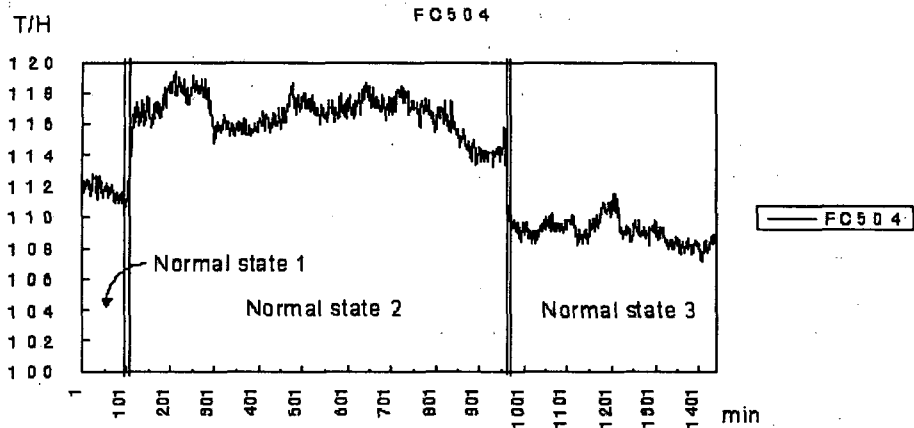


Fig. 1. Labelling example of the sensor data in main drum.

Linear PLS, Neural Network PLS, Proposed NLPLS의 결과를 비교해 본 결과 다음표와 같았다.

Table 1. Results comparison using one PC

	Linear PLS	Neural Network PLS	Proposed NLPLS
MSEP	0.6556	0.1980	0.0180
Infinite Norm	2.3400	0.8554	0.3020

4. 결론 및 토론

본 연구에서는 보다 빠르고 정확한 가스공정 이상진단을 위해 RBF(radial basis function)와 on-line SOFM(self-organizing feature map)을 적용하여 다변량 통계기법의 하나인 부분최소자승법(non-linear partial least square)을 이용한 모델을 제안하였다. 그리고 제안된 모델의 성능을 비교하기 위하여 제안된 모델을 이용하여 BTX공정에서 중요한 역할을 하는 스팀 생산 보일러 공정에 적용시켜 본 결과 위의 표에서 보이듯이 linear PLS, NNPLS에 비해 상당히 우수한 성능을 지님을 확인할 수 있었다. 이처럼 제안된 모델을 가지고 가스공정의 이상진단에 적용시킨다면 효과적인 관리가 이루어질 것이 예상된다.

5. 감사

본 연구는 교육부를 통한 두뇌한국 21 사업과 국가지정연구실 사업 지원금에

의한 것입니다.

6. 참고문헌

1. Geladi, P., B. R. Kowalski, "Partial Least-Squares Regression : A Tutorial", *Anal. Chim. Acta*, 185, 1(1986a)
2. T. Kohonen, "Self-Organization and Associative Memory", 2nd Ed. Berlin: Springer-Verlag, (1987)
3. T. Hastie, W. Stuetzle, "Principal Curves", *J. of Ame. Sta. Ass.*, 84, 502(1989)
4. Simon Haykin, "Neural Networks : A Comprehensive Foundation", 2nd Ed., Prentice-Hall, (1999)
5. G. Baffi, E. B. Martin, A. J. Morris, "Non-linear projection to latent structures revisited(the neural network PLS algorithm)", *Computers and Chemical Engineering*, 23, pp. 1293-1307(1999)
5. 정신호, "비선형 주성분 해석을 이용한 화학공정의 이상감지에 관한 연구", 석사학위논문, 서울대학교 응용화학부(2000)
6. 손정현, "비선형 부분최소자승법을 이용한 화학공정의 추론모델개발에 관한 연구", 석사학위논문, 서울대학교 응용화학부(2001)