

지도학습기법을 이용한 비선형 다변량 공정의 비정상 상태 탐지

손영태 · 윤덕균[†]

한양대학교 산업공학과

Abnormality Detection to Non-linear Multivariate Process Using Supervised Learning Methods

Young-Tae Son · Deok-Kyun Yun

Department of Industrial Engineering, Hanyang University, Seoul 133-791, Korea

Principal Component Analysis (PCA) reduces the dimensionality of the process by creating a new set of variables, Principal components (PCs), which attempt to reflect the true underlying process dimension. However, for highly nonlinear processes, this form of monitoring may not be efficient since the process dimensionality can't be represented by a small number of PCs. Examples include the process of semiconductors, pharmaceuticals and chemicals. Nonlinear correlated process variables can be reduced to a set of nonlinear principal components, through the application of Kernel Principal Component Analysis (KPCA). Support Vector Data Description (SVDD) which has roots in a supervised learning theory is a training algorithm based on structural risk minimization. Its control limit does not depend on the distribution, but adapts to the real data. So, in this paper proposes a non-linear process monitoring technique based on supervised learning methods and KPCA. Through simulated examples, it has been shown that the proposed monitoring chart is more effective than T^2 chart for nonlinear processes.

Keyword: supervised learning theory, non-linear multivariate process monitoring technique, support vector data description, kernel principal component analysis

1. 서론

무한경쟁시대에서 살아남기 위한 기업의 차별화된 국제 경쟁력은 여러 가지가 있겠지만 제조 공정에서 제품의 품질 특성이 가장 중요할 것인 가장 중요한 것은 정해진 공정 설계 규격에 얼마나 일정할 수 있는냐하는 균질성(homogeneity)이라고 할 수 있다. 이처럼 제조공정에서는 제조의 목표가 되는 설계 품질에 합치되는 양질의 제품을 만들어 내는 일이 주요 목표 중의 하나이다.

오늘날 센서 및 측정 기술의 발달에 따라 공정으로부터 얻어지는 자료의 양이 많아지고 시간 간격도 짧아지고 있다. 이러한 매초 또는 매분 단위로 수집되고 있는 많은 다변량 자료를 어떻게 처리하여 공정의 상태를 모니터링하고 이상을 조기에 발견(fault detection), 진단(diagnosis)하여 고품질의 제품을 생산하는가 하는 것이 오래전부터 큰 관심거리로 대두 되었다. 온라인 공정 모니터링은 상관관계가 존재하는 다수의 공정변수들을 동시에 실시간 관리하는 방법이다. 다변량 관리도는 공정변수들 간의 상관관계를 고려하기 때문에 서로 상관관계

[†]연락처 : 윤덕균 교수, 133-791 서울시 성동구 행당동 17번지 공업센터 7층 한양대학교 산업공학과,

Fax : 02-2220-0577, E-mail : ydkshs@hanyang.ac.kr

투고일(2010년 11월 11일), 심사일(1차 : 2010년 12월 30일, 2차 : 2011년 01월 09일), 게재확정일(2011년 01월 14일).

가 존재하는 경우 개개의 공정변수들을 단변량 관리도를 이용하여 독립적으로 관리하는 것보다 공정 변동에 민감하다는 장점이 있다.

일반적인 다변량 정규성과 같은 특정 분포에 대한 가정을 가지는 모수적 방법과 달리 Sun and Tsung(2003)은 특정 분포에 대한 가정이 필요치 않는 SVDD(Support Vector Data Description) 기반의 비모수 공정 모니터링 방법을 제안하였다. Kumar *et al.*(2006)은 표준화된 거리를 이용한 개선된 SVDD 알고리즘을 제안하였으며, Sukchotrat *et al.*(2010)은 부스트랩을 이용한 개선된 관리한계선 추정 방법을 제안하였다. Gani *et al.*(2010)은 실제 냉장고 제조업체의 공정 모니터링에 적용하였고, 시뮬레이션을 통해 T^2 차트와의 성능을 비교 수행하였다. 이처럼 SVDD를 이용한 온라인 공정 모니터링 기법은 이미 많은 연구자들에 의해 그 실효성이 검증된 방법이다.

대량의 공정변수 정보들이 온라인으로 수집되며 입력과 출력 변수 사이에 강한 비선형성(non-linearity)을 가지는 반도체 식각(etch) 공정, 자동차의 도장(painting) 공정 등에 기존의 선형성에 기반한 PCA(principal component analysis)와 PLS(partial least squares) 등을 적용하는 것은 효율적이지 못하다.

Vanrolleghem *et al.*(2004)는 비선형 공정에 비선형 주성분분석인 KPCA(kernel principal component analysis) 기반의 T^2 , SPE 관리도를 제안하였고, Yoo *et al.*(2006)는 MEWMA(multivariate exponentially weighted moving average)를 적용 하였다.

본 연구의 목적은 입력과 출력 변수간의 비선형 관계를 무한차수의 커널함수(Kernel Function)를 이용하여 고차원의 특성공간(Feature Space)에 사상(mapping)하는 KPCA를 수행함으로써 공정 데이터의 비선형성 문제를 해결하고, 학습 이론 분야에서 일반화 성능이 우수한 것으로 알려진 SVDD를 이용한 공정 모니터링 모델을 제안하며 실험을 통해 신뢰성을 검증하고자 한다.

본 논문의 구성은 다음과 같다. 본 논문의 제 2장에서는 KPCA, SVDD 이론에 대해서 기술하며, 제 3장에서는 제안하는 모델에 대해서 기술한다. 제 4장에서는 실험 결과에 대해 언급하며, 마지막으로 결론 및 추후 연구를 제 5장에서 제시한다.

2. 이론적 배경

2.1 Kernel PCA

1991년 Kramer는 auto-associative neural networks(신경망) 기반 비선형 주성분 분석기법을 제시했지만 이는 총 5개의 layer로 구성되어 있기 때문에 학습이 매우 어렵고 각 layer의 노드 수를 결정하는데 어려움이 있다. 1996년 Dong과 McAvoy가 principal curve와 신경망을 이용한 방법이 제안되었으나 이 방법 또한 비선형 최적화 문제를 풀어야 하고 신경망을 학습하기 전에 미리 주성분의 수를 결정해야 하는 단점이 있다.

KPCA는 PCA의 확장된 개념으로 Scholkopf(1998)가 제안한

비선형 주성분 분석법으로 데이터의 비선형성을 해결하기 위해 입력공간(input space)이 아닌 특성 공간에서 PCA를 수행한다는 차이점이 있다.

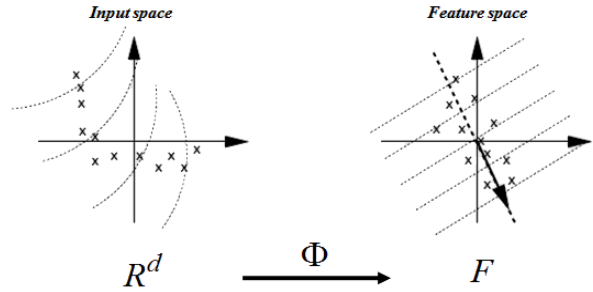


Figure 1. Basic Idea of KPCA

M 개의 공정 벡터 $x_i \in R^d$, $i = 1, 2, \dots, M$ 으로부터 고차원의 특성공간으로 사상하는 비선형 커널함수를 ϕ 라 하면 식 (1)과 같이 표현할 수 있다.

$$\Phi : R^d \rightarrow F, x \mapsto \Phi(x) \quad (1)$$

여기서 $\Phi(x)$ 는 x 가 고차원으로 변환된 특성벡터이며, R^d (d 차원)를 입력공간이라고 하며, F 는 임의의 고차원 공간으로 특성공간이라 한다. $\Phi(x)$ 로 사상된 특성공간의 특성벡터는 커널 트릭(kernel trick)을 이용하여 간단하게 구할 수 있다.

특성공간에 사상된 벡터를 $\Phi(x_1), \dots, \Phi(x_M) \in F$, $(\sum_{i=1}^M \Phi(x_i) = 0)$ 이라 가정하면 공분산 행렬 C^F 는 식 (2)와 같다.

$$C^F = \frac{1}{M} \sum_{j=1}^M \Phi(x_j) \Phi(x_j)^T \quad (2)$$

이 공분산 행렬의 양의 고유치 λ 에 대해 식 (3)을 만족하는 고유펙터 v 를 구함으로써 특성공간에서 PCA를 수행할 수 있다.

$$\lambda v = C^F v \quad (3)$$

하지만 무한의 차원을 가질 수 있는 ϕ 함수로 사상된 특성공간에서의 $\Phi(x)$ 와 v 가 고차원의 특성공간에 존재하기 때문에 계산비용의 증가를 초래한다. 계산량이 감당할 수 없을 정도로 커지므로 이 문제는 커널 트릭을 이용한다. 커널 함수 $K(x_i, x_j)$ 는 식 (4)과 같다.

$$K(x_i, x_j) = (\Phi(x_i) \cdot \Phi(x_j)) \quad (4)$$

이는 특성공간의 사상 Φ 를 적절히 선택하면 특성공간에서의 내적(inner product)은 입력공간에서의 커널과 동일한 것으로

고차원의 공분산 행렬 대신에 내적 행렬을 이용해서 수행된다.

식 (3)의 우변은 $C^F v = \frac{1}{M} \sum_{j=1}^M (\Phi(x_j) \cdot v) \Phi(x_j)^T$ 와 같고 양변에 $\Phi(x_k)$ 를 곱하면 식 (5)과 같이 표현할 수 있다.

$$\lambda(\Phi(x_k) \cdot v) = (\Phi(x_k) \cdot C^F v) \quad (5)$$

v 를 $\Phi(x_1), \Phi(x_2), \dots, \Phi(x_M)$ 의 선형결합(linear combination) 이라 가정하고, 계수(coeffcient) $\alpha_i (i=1, 2, \dots, M)$ 이라 하면 v 는 식 (6)과 같이 표현할 수 있다.

$$v = \sum_{i=1}^M \alpha_i \Phi(x_i) \quad (6)$$

식 (5)에 식 (6)을 대입하면 아래와 같은 식 (7)을 유도할 수 있다.

$$\begin{aligned} \lambda \sum_{i=1}^M \alpha_i (\Phi(x_k) \cdot \Phi(x_i)) \\ = \frac{1}{M} \sum_{i=1}^M \alpha_i \left(\sum_{j=1}^M (\Phi(x_k) \cdot (\Phi(x_j) \cdot (\Phi(x_j) \cdot \Phi(x_i)))) \right) \\ \text{for all } k=1, 2, \dots, M \end{aligned} \quad (7)$$

커널 행렬($M \times M$) K 를 $\Phi(x_i)$ 와 $\Phi(x_j)$ 의 내적으로 정의하면 식 (8)과 같고,

$$K = (\Phi(x_i) \cdot \Phi(x_j)) \quad (8)$$

이로부터 식 (7)은 아래와 같은 식 (9)로 표현할 수 있다.

$$M \lambda K \alpha = K^2 \alpha \quad (9)$$

식 (9)의 양변을 K 로 나누면 식 (10)과 같고,

$$M \lambda \alpha = K \alpha \quad (10)$$

$\alpha = [\alpha_1, \alpha_2, \dots, \alpha_M]$ 는 고유치 문제로 풀 수 있다. 이 값들이 식 (3)과 같은 방법으로 고유벡터의 크기가 1이라는 정의를 만족시키기 위해 식 (11)과 같이 정규화(normalization)를 한다.

$$\begin{aligned} (v_k \cdot v_k) &= 1 \text{ for all } k=1, 2, \dots, p, \\ 1 &= \sum_{i=1}^M \sum_{j=1}^M \alpha_i^k \alpha_j^k (\Phi(x_i) \cdot \Phi(x_j)) \\ &= (\alpha_k \cdot K \alpha_k) = \lambda (\alpha_k \cdot \alpha_k) \end{aligned} \quad (11)$$

마찬가지로 커널 행렬 또한 데이터의 평균을 0으로 보정하기 위해 스케일링된 커널행렬 식 (12)를 이용한다.

$$\tilde{K} = K - \frac{1}{M} K \mathbf{1}_M - \mathbf{1}_M K + \frac{1}{M} \mathbf{1}_M \mathbf{1}_M^T \quad (12)$$

where

$$\mathbf{1}_M = \frac{1}{M} \begin{bmatrix} 1 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 1 \end{bmatrix} \in R^{M \times M}$$

새로운 샘플 x 에 대한 k -번째 주성분 t_k 은 식 (13)과 같이 구할 수 있다.

$$t_k = (V^k \cdot \Phi(x)) = \sum_{i=1}^M \alpha_i^k (\Phi(x_i) \cdot \Phi(x)) \quad (13)$$

커널 함수는 Mercer의 조건을 만족하는 함수들이 사용가능하다고 알려져 있으며 대표적인 커널 함수로는 <Table 1>과 같다(Vapnik, 1998).

Table 1. Typical Example of Kernel Function

Kernel function	$K(x, y)$
RBF	$K(x, y) = \exp\left(-\frac{\ x-y\ ^2}{\sigma}\right)$
Polynomial	$K(x, y) = (x, y)^d$
Sigmoid	$K(x, y) = \tanh(\beta_0(x, y) + \beta_1)$

2.2 SVDD(Support Vector Data Description)

기계학습(Machine Learning)의 대표적인 한 가지 방법인 지도 학습 이론(supervised learning theory)은 기존의 알려진 데이터를 기반으로 학습하여 새로운 데이터를 적절한 클래스로 분류(classification)하는 문제를 다루고 있다. 의사결정나무(decision tree), 신경망(neural networks), SVM(Support Vector Machine) 등이 이를 위해 일반적으로 사용되는 방법들이다.

SVM의 한 종류인 SVDD(Tax, 1999)는 구조적 위험 최소화(structural risk minimization) 원리에 기반하고 있는 단일 클래스 분류 문제(one-class classification problem)에 우수한 일반화 성능을 보장하는 학습 알고리즘으로 비정상상태 탐지(abnormality detection), 군집화(clustering) 등과 같은 다양한 분야에서 활용되고 있다. 단일 클래스 분류 문제는 분류 대상이 되는 하나의 학습 클래스(target class)에 속한 데이터만을 이용하여 학습(training)을 수행하는데 일반적으로 학습과정에서 정상 클래스(normal class) 위주의 데이터만을 다루게 되고, 학습결과로 얻은 결정 함수(decision function)는 주어진 시험 데이터가 정상상태 범주를 벗어나는 비정상 클래스(abnormal class)에 속하는지를 판단하게 된다.

이와 같은 지도 학습 기반의 분류문제는 학습데이터에 기반하는데 일반적으로 학습 데이터는 positive data(정상상태의 data)와 negative data(비정상상태의 data)로 구성된다. 다변량 공정의 비정상 상태 탐지문제에서는 공정이 정상 상태일 때 정상상태의 데이터에 비해 비정상 상태의 데이터가 상대적으로 부

족하기 때문에 negative 데이터의 구성에 어려움이 존재한다. 일반적인 분류 모델이 이진 분류 혹은 다중 분류 문제를 다룰 수 있지만 SVDD는 이처럼 negative data가 상대적으로 적게 존재하는 다변량 공정의 비정상 상태 탐지 문제를 효과적으로 예측할 수 있다.

SVDD의 기본 아이디어는 정상 클래스에 속하는 주어진 학습 데이터를 특성공간에서 정의되는 구(B)로 근사하는데 이는 학습 데이터가 상당수 포함할 수 있으면서 체적을 가장 적게 하는 중심 a와 구의 반경이 R인 최소한의 구(hypersphere)로 구성된다.

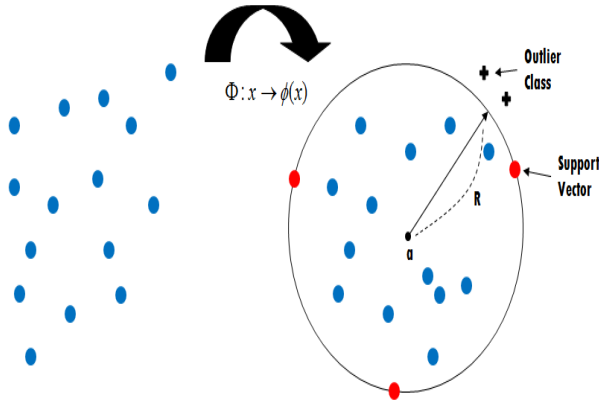


Figure 2. The Concept of SVDD Classifier

이를 위해 특성공간에 존재하는 학습 데이터 $X = \Phi(x_i) | \Phi(x) \in F, i = 1, 2, \dots, N$ 를 중심 a와 반경이 R인 구를 이용해서 정상 클래스의 영역을 표현하는 것을 고려하면 학습 데이터 $\Phi(x_i)$ 와 중심(a) 사이의 거리는 R이상이 되도록 하고 이 거리가 R을 초과하는 경우 적절한 벌점(penalty)을 부과하는 전략을 사용하면 되는데 이는 다음과 같은 최적화 문제 식 (14)와 같이 표현할 수 있다.

$$\min R^2 + C \sum_{i=1}^N \xi_i \quad (14)$$

$$s.t. \|\Phi(x_i) - a\|^2 \leq R^2, \xi_i \geq 0, \forall i$$

여기에서, ξ_i 는 i -번째 학습 데이터 $\Phi(x_i)$ 가 구를 벗어나는 벌점이며, C는 구의 반경(R)과 벌점항의 상대적 중요성을 조정하는 상수(trade-off constant)이다.

위 문제에 대응하는 쌍대 문제(dual problem)로 바꾸기 위해 각 제약조건에 하나씩의 라그랑제 승수(Lagrange multiplier), 즉, 쌍대 변수(dual variables)를 도입하여 라그랑제 함수(Lagrange function)로 표현하면 식 (15)를 얻는다.

$$L = R^2 + C \sum_{i=1}^N \xi_i \quad (15)$$

$$- \sum_{i=1}^N \alpha_i R^2 + \xi_i - (\|\Phi(x_i) - a\|^2) - \sum_{i=1}^N \gamma_i \xi_i$$

where

$$\alpha_i \geq 0, \gamma_i \geq 0$$

식 (15)에서 α_i, γ_i 는 lagrangian multiplier를 나타낸다. L을 R, a, ξ_i 에 대해서 각각 편미분하여 0으로 놓으면 식 (16)~식 (18)을 얻는다.

$$\frac{\partial L}{\partial R} = 0 : \sum_i \alpha_i = 1 \quad (16)$$

$$\frac{\partial L}{\partial a} = 0 : a = \frac{\sum_i \alpha_i \Phi(x_i)}{\sum_i \alpha_i} = \sum_i \alpha_i \Phi(x_i) \quad (17)$$

$$\frac{\partial L}{\partial \xi_i} = 0 : \gamma_i = C - \alpha_i, \forall_i \Rightarrow 0 \leq \alpha_i \leq C \quad (18)$$

R, a, ξ_i 를 소거하기 위해서 위의 조건을 다시 라그랑제 함수에 대입하면 식 (19)와 같이 표현할 수 있다.

$$\max \sum_i \alpha_i \Phi(x_i) \cdot \Phi(x_i) - \sum_i \sum_j \alpha_i \alpha_j K(x_i, x_j)$$

$$s.t. \sum_i \alpha_i = 1, 0 \leq \alpha_i \leq C, \forall_i$$

$$where K(x_i, x_j) = (\Phi(x_i) \cdot \Phi(x_j)) \quad (19)$$

식 (19)는 KKT 조건(Karush Kuhn Tucker Complementary problem)으로부터 다음 식 (20)이 성립한다.

$$\alpha_i (R^2 + \xi_i - \|\Phi(x_i) - a\|^2) = 0 \quad (20)$$

KKT 조건으로부터 $\|\Phi(x_i) - a\|^2 < R^2$ 이면 $\alpha_i = 0$ 이 되고, $\|\Phi(x_i) - a\|^2 = R^2$ 이면 $0 \leq \alpha_i \leq C$ 이며 이 $\Phi(x_i)$ 는 서포트 벡터이다. $\|\Phi(x_i) - a\|^2 > R^2$ 이면 $\alpha_i = C$ 이며 이 $\Phi(x_i)$ 는 아웃라이어이다.

dual problem 식 (19)로부터 구의 반경(R)과 중심(a)를 구할 수 있다. 특성 공간에서의 최적의 중심(a)는 식 (17)을 통해 구할 수 있고, 구의 반경(R)은 식 (20)을 이용해서 구할 수 있다. 따라서 임의의 서포트 벡터 $\Phi(x_i)$ 로부터 구의 중심(a)에서 구의 반경(R)의 거리는 식 (21)과 같이 계산 된다.

$$R_F^2 = \|\Phi(x_i) - a\|^2 \quad (21)$$

$$= K(x, x) - 2 \sum_i \alpha_i K(x_i, x) + \sum_i \sum_j \alpha_i \alpha_j K(x_i, x_j)$$

$$where K(x_i, x_j) = (\Phi(x_i) \cdot \Phi(x_j))$$

또한, 새로운 시험 데이터(z)와 센터(a)까지의 거리는 식 (22)와 같이 계산 된다.

$$D_F^2 = \|\Phi(z) - a\|^2 \quad (22)$$

$$= K(z, z) - 2\sum_i \alpha_i K(x_i, z) + \sum_i \sum_j \alpha_i \alpha_j K(x_i, x_j)$$

3. 제안 기법

기존의 전통적인 T^2 , MCUSUM, MEWMA 관리도는 기본적으로 변수들 간의 선형성과 정규성을 가정한 방법들이다. 하지만 실제 문제에서 모집단의 분포를 정확하게 반영하는 것이 불가능할 뿐만 아니라 이와 같은 가정을 완벽하게 만족하기는 어렵기 때문에 신뢰할 수 없는 결과를 가져오곤 한다.

따라서 본 연구에서는 입-출력 시스템 간 강한 비선형 관계를 가지는 공정 변수를 KPCA를 통해 해결하고, 특정분포의 가정이 필요 없는 SVDD를 이용한 공정 모니터링 모델을 제안한다.

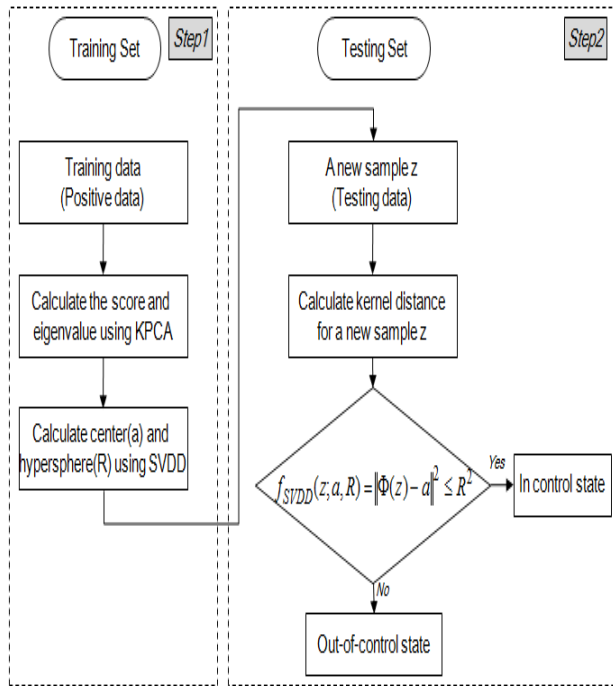


Figure 3. Conceptual Framework of the Proposed Model

이는 <Figure 3>와 같은 절차를 통해 수행되는데 실행과정을 정리하면 다음과 같다.

Step 1: 먼저 정상 상태의 학습 데이터로부터 공정변수의 비선형성을 해결하기 위해 KPCA를 수행한다. 이를 통해 얻은 주성분을 가지고 SVDD를 수행하여 구의 경계(R)와 센터(a)를 얻는다.

Step 2: 학습결과로 얻어진 결정 함수로부터 새로운 시험 데이터(testing data)가 진입했을 때 $\|\Phi(z) - a\|^2 \leq R^2$ 이면 정상 상태로 그렇지 않으면 비정상 상태로 판정하는 전략을 취한다.

4. 수치 실험 및 결과

제안하는 공정 모니터링 모델의 성능을 검증하기 위해 다음과 같은 비선형 2×2 System의 공정 변수를 고려한다.

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} (u_1(t) + u_2(t))^2 \\ (u_1(t) - ku_2(t))^2 \end{bmatrix} \quad (23)$$

여기에서, y_1, y_2 는 출력 변수, u_1, u_2 는 입력변수이며 k 는 2이고 두 종류의 입력은 아래와 같은 Gaussian noise distribution을 통해 얻은 랜덤변수라 가정한다.

$$\mu_0 = [0 \ 0], \quad \Sigma_0 = \begin{bmatrix} 1 & .3 \\ .3 & 1 \end{bmatrix} \quad (24)$$

위의 Gaussian noise distribution을 따르는 두 입력 변수와 출력 변수로부터 100개의 학습 데이터를 구성하고 분산의 변화는 고려하지 않은 상태에서 평균의 변화에 따른 랜덤하게 생성된 100개의 시험 데이터를 생성하였다. 또한, 모니터링 결과를 예측하는 모델의 정확도를 평가하기 위해 민감도(sensitivity)와 특이도(specificity)를 이용한다.

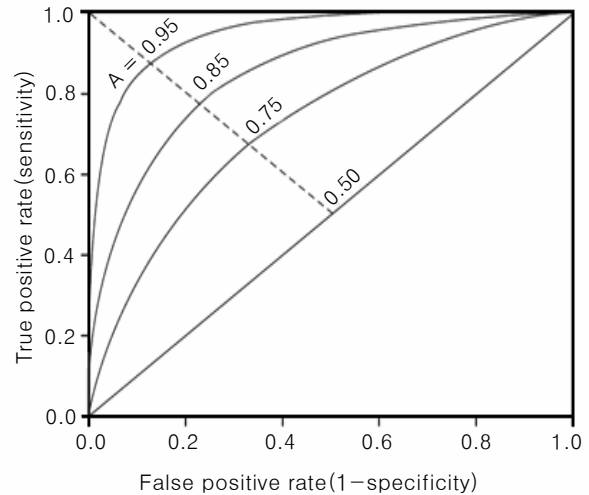


Figure 4. Example of a ROC Curve

민감도는 정상상태를 정상상태로 판단하는 비율이고 특이도는 비정상 상태를 비정상 상태로 판정하는 비율이다. 따라서 민감도와 특이도의 비율이 높을수록 모델의 성능이 좋다고 할 수 있다. 판정의 정확도는 <Figure 4>와 같은 Receiver Operating Characteristics(ROC) curve를 사용하는데 1-특이도(x축)와 민감도(y축)를 그래프로 표현한 것으로 ROC curve 아래의 면적(area under the ROC curve, AUC)을 통해 확인할 수 있다.

일반적으로 AUC의 면적의 수치에 따른 분류 정확도는 <Table 2>와 같은 성능을 가짐을 의미한다.

Table 2. Using the Area under the ROC Curve to Interpret Sensitivity and Specificity

AUC value	Interpretation of the AUC
($AUC = 0.5$)	non predictive
($0.5 < AUC \leq 0.7$)	less predictive
($0.7 < AUC \leq 0.9$)	highly predictive
($AUC = 1.0$)	perfect predictive

즉, 그려진 curve가 좌측 위쪽 모서리에 가까울수록 좀 더 정확한 모델이라고 할 수 있는데, 이는 정상 상태를 정상 상태로 판단하는 비율(민감도)이 높고 비정상 상태를 정상 상태로 판단하는 비율(FPR : false positive rate)이 낮음을 의미한다.

본 논문에서는 식 (25)와 같이 정의되는 가우시안 커널을 고려하며 KPCA 분석에서는 커널 함수의 파라미터 σ 값은 3을 이용한다.

$$K(x, y) = \exp\left(-\frac{\|x - y\|^2}{\sigma}\right) \quad (25)$$

또한, SVDD 분석에서는 조정 상수 C는 0.05, 커널 함수 σ 값은 2로 고정하였고 MATLAB dd_tools 1.7.3(Tax, 2009)를 이용해 분석하였다.

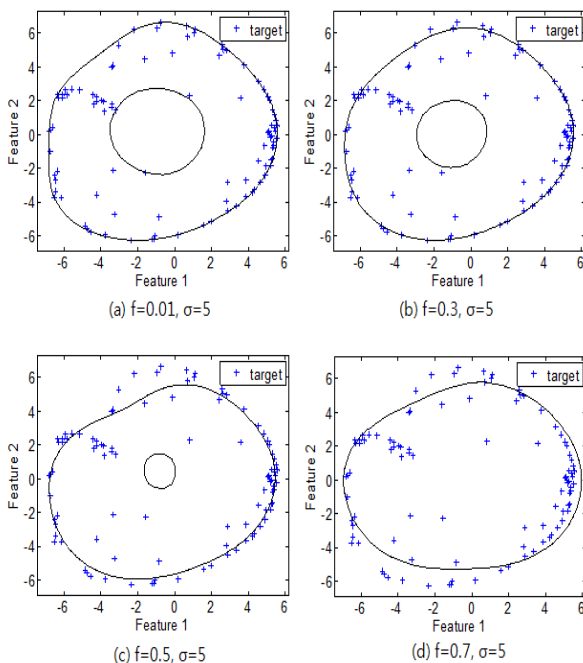


Figure 5. Change of Hypersphere Boundaries with Parameter Value(f)

식 (14)에서 언급했던 조정 상수 C와 커널 함수 σ 값은 결정 함수와 관리 한계의 선정에 영향을 준다. 결과적으로, 두 파라미터의 결정은 결정함수의 성능과 직결되는 문제로 성능관점에서 상당히 중요한 문제이다. <Figure 5>에서 보는바와 같이

C값이 클 경우 구의 경계가 작아짐에 따라 제 1종 오류(type 1 error : 정상상태를 비정상 상태로 판정하는 오류)는 증가하는 반면, C값이 작은 경우 커진 구의 경계로 인해 제 2종 오류(type 2 error : 비정상 상태를 정상상태로 판정하는 오류)가 증가하는 오분류(misclassification) 가능성이 커진다.

또한, <Figure 6>은 커널 함수 σ 값에 따른 구의 경계 변화를 보여주고 있다. 만약 σ 값이 작다면 각 데이터들 간의 거리의 SVDD에 대한 영향력이 커져 데이터가 존재하는 영역을 조밀하게 표현하며, 반대로, σ 가 크다면 각 데이터들 간의 거리의 SVDD에 대한 영향력이 작아져 구에 가까운 영역을 얻게 된다.

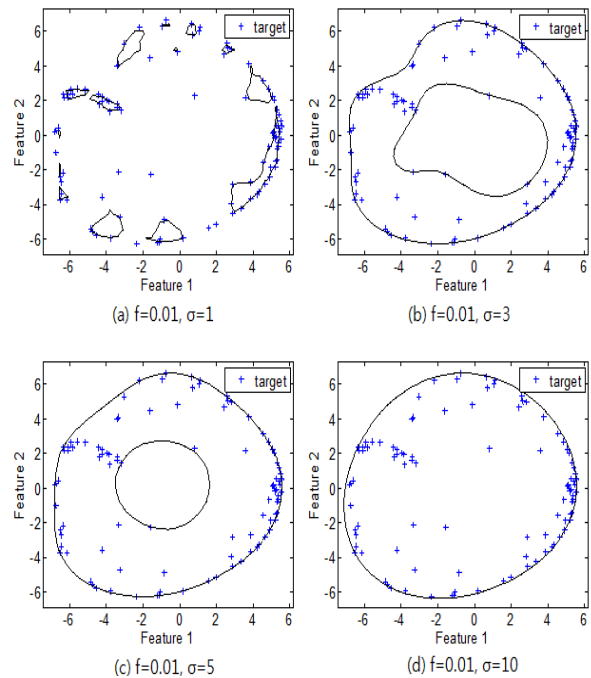


Figure 6. Change of Hypersphere Boundaries with Parameter Value(σ)

본 실험에서는 비선형 공정의 일반적인 상황을 고려하여 기존에 제시된 T^2 관리도와 본 연구에서 제안하는 SVDD 기반의 관리도의 성능 평가 비교를 위해 평균의 변화(Mean shift)에 따른 비 정상상태 데이터를 동일한 조건에서 생성하였고 그에 따른 AUC와 FPR의 변화는 <Table 3>과 같다.

평균의 크기를 순차적으로 변화시켰을 때 <Table 3>에서 보는바와 같이 제안하는 방법이 AUC(분류 정확도)가 높고, FPR(비정상 상태를 정상 상태로 판정하는 오류)가 낮음을 알 수 있다. FPR이 높다는 것은 공정에 이상이 존재함에도 불구하고 생산을 진행시켜 불량품을 양산하는 결과를 초래하여 그에 따른 손실비용을 발생시킬 수 있음을 의미한다. 현재의 공정이 고가의 제품을 소량 생산하는 공정으로의 변화 추세에 따라 기업의 입장에서는 설계 품질과 합치되는 양질의 제품을 양산하는 일은 매우 중요한 문제이다. 이런 점에서 제안하는 기법이 비선형 공정의 공정모니터링에서 유용하게 적용될 수 있을 것이다.

Table 3. Empirical Results of FPR and AUC between T^2 Chart and the SVDD based Chart

Mean shift	T^2 Chart ($\alpha = 0.05$)		SVDD based Chart ($f = 0.05, \sigma = 2$)	
	AUC	FPR	AUC	FPR
0.2	0.9369	0.37	0.9863	0.14
0.4	0.9158	0.54	0.9833	0.13
0.6	0.9100	0.58	0.9825	0.19
0.8	0.9028	0.58	0.9762	0.27
1.0	0.8911	0.63	0.9690	0.35
2.0	0.8313	0.87	0.9411	0.71

5. 결론 및 추후 연구

일반적인 연속 생산 공정과 달리 반도체, 화학, 제약 공정과 같은 고부가가치 제품을 다품종 소량 생산하는 배치 공정의 효율적인 관리를 위해서는 온라인 공정 모니터링을 통해서 공정의 이상을 조기에 발견하여 제품의 품질을 균일하게 제어하는 것이 무엇보다 중요하다.

따라서 본 논문은 고도의 비선형성을 가지는 공정의 이상상태를 감지하기 위한 새로운 비선형 다변량 공정 모니터링 모델을 제안하였고 실험 데이터에 적용해 보았으며, 그 결과 제안하는 학습 이론을 이용해 구현한 모델이 효과적으로 공정의 이상을 감지함을 확인하였다.

제안된 SVDD 기법은 특정분포에 대한 가정을 필요치 않으며 결정함수를 구성함에 있어 구의 경계(R)과 센터(a)를 제외한 학습데이터는 계산에 포함시키지 않기 때문에 계산 시간이 빠르다. 따라서 현실적으로 얻어지는 많은 공정변수와 데이터를 고려할 수 있을 뿐만 아니라 일반화 성능 즉, 새로운 데이터의 분류에 있어서 높은 정확도를 가지는 것으로 알려져 있어 제안된 방법이 실제 생산 공정에 적용될 수 있을 것으로 기대된다.

본 연구에서 제안하는 학습이론을 이용한 공정 모니터링 모델은 기본적으로 공정이 관리 상태에 있을 때 얻어진 과거의 학습데이터에 기반하여 모델을 만들어 모니터링을 수행하므로 학습데이터의 구성 및 그 특성에 따라 그 성능에 차이를 보이게 된다.

또한, 앞서 언급했던 조정 상수 f 와 σ 의 선택에 따라서도 그 성능에 차이를 보이게 되며 위 두 가지 모수의 적절한 선택 방법을 추후 연구과제로 제시한다.

참고문헌

- Croiser, R. B. (1988), Multivariate Generations of Cumulative Sum Quality Control Schemes, *Technometrics*, 30(3), 291-303.
- Gani, W., Taleb, H., and Limam, M. (2010), An assessment of the kernel-distance-based multivariate control chart through an industrial application, *Quality and Reliability Engineering International*, Early View.
- Kumar, S., Choudhary, A. K., Kumar, M., Shankar, R., and Tiwari, M. K. (2006), Kernel distance-based robust support vector methods and its application in developing a robust K-chart, *International Journal of Production Research*, 44(1), 77-96.
- Lee, J. M., Yoo, C. K., Choi, S. W., Vanrolleghem, P. A., and Lee, I. B. (2004), Nonlinear process monitoring using kernel principal component analysis, *Chemical Engineering Science*, 59, 223-234.
- Lowry, C. A., Woodall, W. H., Champ, C. W., Rigdon, S. E. (1992), A Multivariate Exponentially Weighted Moving Average Control Chart, *Technometrics*, 34(1), 46-53.
- Montgomery, D. C. (2005), *Introduction to Statistical Quality Control*, Fifth edition. Wiley, New York, NY.
- Pignatiello, J. J., Jr. and Runger, G. C. (1990), Comparison of multivariate CUSUM charts, *Journal of Quality Technology*, 22(3), 173-186.
- Phillipotts, D. (2007), *Nonlinear fault detection and diagnosis using Kernel based techniques applied to a pilot distillation column*, Master's Dissertation, University of Pretoria, South Africa.
- Sukchotrat, T., Kim, S. B., and Tsung, F. (2010), One-class classification-based control charts for multivariate process monitoring, *IIE Transactions*, 42(2), 107-120.
- Schölkopf, B., Smola, A. J., and Müller, K. (1998), Nonlinear component analysis as a kernel eigenvalue problem, *Neural Computation*, 10(5), 1299-1399.
- SUN, R. and TSUNG, F. (2003), A Kernel-distance-based multivariate control charts using support vector methods, *International Journal of Production Research*, 41(13), 2975-2989.
- Tax, D. M. J. and Duin, R. P. W. (2004), Support vector data description, *Machine Learning*, 54(1), 45-66.
- Vapnik, V. N. (1998), *Statistical Learning Theory*, Wiley, New York, NY.
- Yoo, C. K. and Lee, I. B. (2006), Nonlinear multivariate filtering and bioprocess monitoring for supervising nonlinear biological processes, *Process Biochemistry*, 41(8), 1854-1863.
- Zhanga, J., Martinb, E. B. and Morrisa, A. J. (1997), Process monitoring using non-linear statistical techniques, *Chemical Engineering Journal*, 67(3), 181-189.



윤덕균

서울대학교 화학공학과 학사
한국과학기술원 산업공학과 석사
한국과학기술원 산업공학과 박사
현재 : 한양대학교 산업공학과 교수
관심분야 : 품질경영, 제약경영이론(TOC)



손영태

한남대학교 비즈니스통계학과 학사
현재 : 한양대학교 산업공학과 석사과정
관심분야 : 품질경영, 품질공학