

# SOM기반 특징 신호 추출 기법을 이용한 불균형 주기 신호의 이상 탐지

김승이<sup>1</sup> · 강지훈<sup>1</sup> · 박종혁<sup>1</sup> · 김성식<sup>1</sup> · 백준걸<sup>1\*</sup>

## Fault Detection of Unbalanced Cycle Signal Data Using SOM-based Feature Signal Extraction Method

Song-Ee Kim · Ji-Hoon Kang · Jonghyuck Park · Sung-Shick Kim · Jun-Geol Baek

### ABSTRACT

In this paper, a feature signal extraction method is proposed in order to enhance the low performance of fault detection caused by unbalanced data which denotes the situations when severe disparity exists between the numbers of class instances. Most of the cyclic signals gathered during the process are recognized as normal, while only a few signals are regarded as fault; the majorities of cyclic signals data are unbalanced data. SOM(Self-Organizing Map)-based feature signal extraction method is considered to fix the adverse effects caused by unbalanced data. The weight neurons, mapped to the every node of SOM grid, are extracted as the feature signals of both class data which are used as a reference data set for fault detection. kNN(k-Nearest Neighbor) and SVM(Support Vector Machine) are considered to make fault detection models with comparisons to Hotelling's  $T^2$  Control Chart, the most widely used method for fault detection. Experiments are conducted by using simulated process signals which resembles the frequent cyclic signals in semiconductor manufacturing.

**Key words** : Feature Signal Extraction, SOM(Self-Organizing Map), Unbalanced Data, Cyclic Signal

### 요약

본 연구는 공정신호가 불균형 데이터인 경우 이상 탐지 알고리즘의 성능 개선을 위한 특징 신호 추출 기법을 제안한다. 불균형 데이터란 범주 구분 문제에서 하나의 범주의 속하는 데이터의 비율이 다른 범주의 데이터에 비해 크게 차이나 이상 탐지 성능이 크게 저하되는 경우를 의미한다. 공정이 운영되는 경우 얻을 수 있는 이상 신호의 수는 정상 신호에 비해 매우 적기에 이러한 문제를 해결하여 이상 탐지 기법을 적용하는 것은 매우 중요하다. 불균형 문제 해결을 위해 SOM(Self-Organizing Map) 알고리즘을 이용하여 각 노드에 대응되는 가중치를 특징 신호로 간주하여 정상 데이터와 이상 데이터의 비율을 맞춘다. 특징 신호 데이터 집단의 이상 탐지를 위해 클래스 분류 기법인 kNN(k-Nearest Neighbor)과 SVM(Support Vector Machine)을 적용하여 이를 공정 신호 이상탐지를 위해 주로 사용하는 Hotelling's  $T^2$  관리도와 성능을 비교한다. 반도체 공정에서 발생하고 알려진 공정 신호를 모사하여 신호 알고리즘 성능의 우수성을 검증한다.

**주요어** : 특징 신호 추출, SOM(Self-Organizing Map), 불균형 데이터, 주기 신호

## 1. 서론

작업자가 직접 관여하던 제조 시스템은 에너지를 전환

시켜주는 기관(engine)의 개발에 따라 기계에 의한 생산 시스템으로 변경되었다. 이는 품질에 영향을 미치는 주요 원인을 작업자의 숙달 정도에서 공정 장비의 상태와 운영 조건에 좌우되도록 변화시켰다. 따라서 공정 운영의 경쟁력을 확보하기 위해서는 공정이 항상성(stationary)을 유지하여 운영될 수 있도록 공정 장비의 상태를 지속적으로 관리하는 것이 중요하다(Park, et al., 2011).

공정 장비의 상태를 관리하기 위해서 다양한 방법들이 연구되었다. 가장 전통적인 방법으로는 단변량 SPC(Statistical

접수일(2012년 3월 22일), 심사일(1차 : 2012년 4월 25일),

게재 확정일(2012년 5월 7일)

<sup>1)</sup> 고려대학교 산업경영공학과

주 저 자 : 김승이

교신저자 : 백준걸

E-mail; jungeol@korea.ac.kr

Process Control) 관리도를 적용하는 것이다(Montgomery, 1996). 제품이 생산된 뒤에 만들어진 제품의 특성치, 예를 들면 제품의 길이, 무게, 농도 등 측정 가능한 변량을 측정하고 이들의 통계적 분포를 계산하여 공정이 항상성을 유지하고 있다고 말할 수 있는 정상범위에 있는지 그렇지 않은지를 판단하는 것이다.

현대에 이르러 IT(Information Technology)의 발달로 장비에서 측정되는 실시간 데이터를 획득하는 것이 가능해졌다. 생산된 제품의 특성치를 이용하는 것보다 실시간으로 발생하는 정보를 활용한 품질관리는 공정 상태를 조기에 진단하여 불량 제품 생산 방식을 가능하게 한다. 이처럼 공정의 진행에 따라 발생하는 정보를 지니는 주기성을 가진 특정한 패턴 신호를 주기 신호(cycle signal)라 부른다(이재현 등, 2007).

주기 신호는 공정이 항상성을 유지하는 경우 동일한 형태로 생성된다. 반면 공정에 변화가 존재하는 경우 다른 신호 형태를 드러내는데 이를 이상(fault)이라 한다. 따라서 공정 주기 신호의 이상을 판단하여 공정을 점검하는 이상 탐지(fault detection)는 공정의 잠재적인 불량요인을 조기에 탐지하여 품질을 향상 시키는데 큰 역할을 차지한다.

주기 신호의 이상을 탐지하는 경우에는 한 시점을 하나의 변수로 간주하고 SPC 관리도를 운영한다. 주기 신호가 획득되는 시간에 대한 영향을 반영 할 수 없기 때문에 CUSUM(Cumulative SUM) 관리도와 EWMA(Exponentially Weighted Moving Average) 관리도가 고려된다(이재현 등, 2007). 두 기법 모두 시간에 대한 영향력을 고려하여 기존의 단변량 SPC 관리도에 비해 탐지력이 증가하는 반면 주기가 길어짐에 따라 관리도의 수가 늘어나 비효율적이다.

주기 신호를 하나의 관리도로 관리하기 위해서 다변량 통계 분석 기법이 이용된다. 대표적인 방법은 Hotelling's  $T^2$  관리도, PCA(Principal Component Analysis), 그리고 PLS(Partial Least Squares)가 있다. 이와 같은 다변량 통계 분석 기법은 여러 개의 공정 변수를 고려하여 프로세스를 관리하는 기법이다(Shi and Jin, 2000). 이 외에도 이상 탐지를 위해서 웨이블릿 기법(이재현 등, 2007), SVM-R(Support Vector Machine-Regression)(박승환 등, 2010), 스플라인 회귀 분석 및 SVM(Support Vector Machine)(Park, et al., 2011) 등이 연구되었다.

기존 연구는 주기 신호의 이상 탐지를 하는데 있어서 뛰어난 효과를 보였다. 하지만 위의 방법들은 이상 주기 신호를 사용하지 않고 이상 탐지를 진행하였거나 이상 주기 신호와 정상 주기 신호의 개수가 상이한 경우 발생하

는 문제점을 고려하지 않았다. 주기 신호는 공정의 상태를 반영하기 때문에 이상 신호를 이용하는 것은 이상 탐지의 정확성을 증가시키는 장점을 가진다. 이상 신호를 활용한다 하더라도 실제의 공정 데이터는 정상 신호의 비율이 높고 이상 신호의 비율이 낮은 형태를 지니는데 이를 반영하지 못하면 이상 탐지 성능을 저하시킨다.

실제 공정에서 정상 신호와 이상 신호 비율이 크게 차이 나듯 범주의 비율이 크게 차이 나는 자료를 불균형 데이터(unbalanced data)라 한다(강필성 등, 2005). 불균형 데이터는 다음과 같은 이유에서 범주 구분, 주기 신호의 경우 이상 탐지에 어려움을 지닌다. 먼저 불균형 데이터의 경우 다수 범주에 속한 데이터의 분포가 이상 범주에 속한 데이터의 분포보다 넓어 이상 데이터의 분류 경계(decision boundary)를 명확히 할 수 없다(강필성 등, 2006). 또 소수 범주의 데이터 정보를 상대적으로 적게 반영하는 정보 불균형의 문제 및 데이터 분포가 겹치는 경우 소수 범주를 인식하기 어려운 문제를 지닌다(Sun, et al., 2007).

데이터 불균형 문제를 해결하기 위한 대표적인 방법은 표본 추출법(sampling)이다. 표본 추출 방법에는 다수 데이터에서 임의적 표본 추출을 하여 소수 범주의 데이터와 균형(balance)을 이루도록 하는 과소 표본 추출(under sampling) 방법과 소수 데이터를 반복적으로 복사하여 다수 데이터와의 균형을 맞추는 과대 표본 추출(over sampling) 방법이 있다(강필성 등, 2006). Japkowicz(2000)는 불균형 데이터에서 표본 추출 기법인 과소 표본 추출 방식과 과대 표본 추출 방식을 다른 기법과의 비교를 통해 두 가지 방법이 데이터 불균형을 해결하는데 있어서 우수함을 증명하였다.

과대 표본 추출에 관한 기존 연구는 SMOTE(Synthetic Minority Over-Sampling Technique)가 있다(Chawla, et al., 2002). SMOTE란 기존 과대 표본 추출 방식처럼 소수 범주의 값을 반복하여 추출하는 방법이 아닌, k-최근접 이웃 알고리즘을 이용하여 소수 범주들과 이웃들 사이에서 보간법을 이용하여 인공 데이터를 생성하는 방식이다. SMOTE를 기존의 부스팅 기법과 결합시킨 SMOTEBoost를 제안하기도 하였다(Chawla, et al., 2003). 또한 오장민과 장병탁(2001)은 불균형 데이터 문제를 해결하기 위해서 퍼셉트론을 기반으로 하여 부스팅 기법을 제안하였다. 과대 표본 추출은 모든 데이터의 정보를 사용할 수 있다는 장점이 있는 반면 데이터의 수를 매우 증가시켜 계산에 소요되는 시간이 증가한다는 단점이 있다.

과소 표본 추출에 관한 연구로는 Laurikkala(2001)에

의해 제안된 방식인 NCL(Neighbourhood Cleaning Rule)이 있다. NCL 방식은 소수 범주 데이터와 유사한 다수 범주 데이터를 제거한 후에 표본 추출을 하여 모델을 개발하는 방식이다. NCL 방식은 두 범주 데이터의 패턴을 뚜렷하게 나타나도록 하며 소수 범주 데이터와 다수 범주 데이터를 구분하는데 있어서 높은 성능을 발휘하였다. 강필성 등(2005)은 두 범주의 데이터 비율이 현저하게 차이가 나는 경우에 기존 분류기 성능에 미치는 영향을 분석하여 문제점을 파악하였으며 SVM 앙상블 기법을 적용하여 과소 표본 추출 방법을 제안하였다. SVM 앙상블 기법은 기존 분류기가 불균형 데이터에서 갖는 단점을 해결하였다. 하지만 과소 표본 추출 방식은 다수 범주에서 샘플링 되는 데이터에 따라 분류기의 성능이 크게 영향을 받을 수 있다는 단점이 있다.

본 연구에서는 주기 신호 데이터의 특징 신호를 추출하여 반도체 공정 불균형 데이터의 이상 탐지를 위한 방안을 제시하고자 한다. 기계학습 알고리즘 중 하나인 SOM(Self-Organizing Map) 알고리즘을 이용하여 기존의 정상 신호와 이상 신호를 기반으로 두 범주를 대표 할 수 있는 특징 신호(feature signal)를 추출한다. SOM 신경망의 뉴런(neuron)값을 기반으로 하여 이상 데이터와 정상 데이터의 비율을 맞춘 뒤 클래스 분류 기법을 적용하여 이상 탐지를 진행한다.

본 연구의 2장에서 불균형 문제를 해결하기 위한 특징 신호 추출 기법을, 3장에서 추출된 특징 신호에 대한 분류 기법을 제안한다. 본 연구를 통해 제안된 이상 탐지 방법을 반도체 공정 신호를 모사한 자료를 이용하여 기존의 방법과의 시뮬레이션 비교 분석 결과를 제시하고 결론을 정리하였다.

## 2. 특징 신호 추출

공정 운영을 시험하기 위해 다양한 환경을 가정하고 운영하는 파일럿 공정(pilot process)을 제외하고 반도체 공정은 관리된(in-control) 상황에서 진행되기에 획득되는 신호는 대부분이 정상 신호 범주에 속하는 불균형 데이터이다. 데이터의 범주 비율의 차이가 크게 나타나는 불균형 데이터는 각 범주를 구분하기에 난점이 있기에 이 비율을 맞추어 주는 전처리과정이 필요한데 이를 특징 신호 추출(feature signal extraction)이라 부른다. 앞서 언급한 것처럼 본 연구는 인공 신경망(artificial neural network)의 일종인 SOM 알고리즘을 이용하여 특징 신호 추출 과정을 진행하였다.

### 2.1 Self-Organizing Map

SOM은 어떤 범주에 속하는지 알려지지 않은 경우 입력 데이터의 유사성을 비교하여 데이터의 위상을 인식 할 수 있도록 하는 인공 신경망으로(Kohonen, 1982) 클러스터링(clustering), 패턴인식 및 분류(pattern recognition and classification), 자료 차원축소(dimension reduction) 및 특징 추출(feature extraction)등 다양한 목적으로 활용되어 좋은 성과를 보이고 있다(Fayyad, et al., 1996). SOM은 그림 1에서 보는 바와 같이 입력층과 출력층, 두 개의 층으로 구성된 단일층 퍼셉트론(single-layer perceptron) 구조를 지닌다. 일반적인 퍼셉트론 구조와 다른 것은 출력층이 격자로 된 N개의 뉴런으로 구성되어 있다는 점이다(한학용, 2009). 그림 1은  $6 \times 6$ 의 그리드 크기를 갖는 총 36개의 뉴런으로 구성된 예제이고 그리드란 SOM을 구성하는 뉴런으로 이루어진 2차원의 격자 틀을 말한다.

SOM은 출력층의 각 뉴런을 모든 입력층의 뉴런과 연결하여 입력 데이터와 같은 차원을 가지는 가중치 벡터를 구성하는데 이는 그림에서 각 뉴런을 연결하는 신경망으로 표시된다. 가중치 벡터 및 뉴런 간의 연결 강도를 갱신하는 과정, 즉 인공 신경망 학습 과정은 경쟁 과정, 협동 과정, 적응 과정으로 구성된다. 경쟁(competition) 과정은 모든 뉴런의 연결 강도와와 거리를 계산하여 뉴런 사이에서 거리가 최소화되는 뉴런을 승자 뉴런으로 하는 경쟁 규칙이다. 협동(cooperation) 과정에서는 출력층의 유사한 특징 패턴에 대하여 지도를 형성한다. 지도를 형성하기 위해 승자 뉴런은 이웃한 뉴런을 정해진 함수에 의하여 결정하고 근접한 뉴런의 연결 강도를 갱신한다. 마지막은 적응(adaptation) 과정이다. 적응 과정에서는 승자 뉴런과 그와 이웃한 뉴런들이 특정한 입력 값에 더욱 민감해지도록 활성 함수를 적용시켜 연결 강도를 갱신한다.

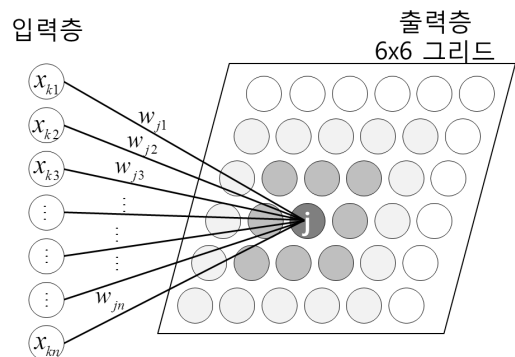


그림 1.  $6 \times 6$  그리드 SOM 예제

이 과정을 통하여 승자 뉴런과 인접한 뉴런들은 멀리 떨어져 있는 뉴런보다 더 강하게 적응 된다(한학용, 2009).

입력되는 데이터는  $n$ 차원이고 해당 데이터가  $K$ 개 존재할 때 입력벡터  $x_k = (x_{k1}, x_{k2}, \dots, x_{kn}), k \in \{1, 2, \dots, K\}$  는 2차원 출력층에 존재하는  $N$ 개의 뉴런에 대응된다. 이때 뉴런과의 대응을 위한 연결 고리로 설정되는 각각의 가중치 벡터는  $w_j = (w_{j1}, w_{j2}, \dots, w_{jn}), j \in \{1, 2, \dots, N\}$  과 같다. 학습을 위해서 0에서 1사이의 정규화 된 값으로 각 뉴런간의 연결 강도를 초기화하고 0에 가까운 임의의 값으로 가중치 벡터를 설정한다. 본 연구에서는 윤재준 등(2010)에서 제시한 것처럼 주기신호의 각 시점별로 샘플링 된 값을 입력벡터로 정의하였다.

이어서 입력 벡터와 모든 가중치와의 유클리드 거리를 계산하여 이 값이 최소가 되는 뉴런을 찾는 과정을 거치는데 이때 최솟값을 갖는 뉴런을 승자 뉴런(winning neuron)이라 부른다.

$$c_k = \arg \min \| x_k - w_j \| \quad (1)$$

식 (1)에 의해 입력 데이터  $x_k$  에 대응되는 승자 뉴런  $c_k$  이 계산된다. 모든 입력 데이터  $x_k, k \in \{1, 2, \dots, K\}$  에 대해 승자 뉴런이 계산되면 승자 뉴런과 해당 뉴런과 가까이 있는 뉴런의 가중치를 업데이트 하게 되는데 식 (2)와 같은 방식으로 진행된다(윤재준 등, 2010).

$$w_j \leftarrow w_j + \alpha_t \cdot h_i(t, c_k) \cdot [x_k - w_j] \quad (2)$$

여기에서  $\alpha_t$  는 시점  $t$  에 대한 학습률을 나타내는 값으로 학습이 진행됨에 따라 값이 점점 줄어들게 된다.  $h_i(t, c_k)$  함수는 입력 벡터  $x_k$  에 대응되는 승자 뉴런  $c_k$  에 인접한 이웃을 나타내주는 함수로  $c_k$  와 가까울수록 큰 값을 가지고 먼 값일수록 작은 값을 가지는데, 이 또한 시간이 지나게 됨에 따라 점차 줄어드는 특성을 지닌다. 입력 벡터와 해당 노드의 가중치의 차의 절대값을 위의 두 값과 곱하여 더해주는 식으로 가중치 벡터를 업데이트 하게 된다. 업데이트는 승자 뉴런 주위의 뉴런을 승자 뉴런과 유사하게 반영 할 수 있으므로 비슷한 입력 벡터를 출력층의 근접한 뉴런에 모을 수 있다는 장점을 지닌다.

모든 데이터에 대하여 업데이트 과정을 진행하고 식 (1)에서 언급한 경쟁 과정과 식 (2)의 협동 및 적응 과정을 반복하여 얻은 최종 승자 뉴런 결과를 이용하여 클러스터링, 출력층의 뉴런의 위치를 이용해 차원을 축소하는 작업 등이 가능하다.

## 2.2 SOM을 이용한 특징 신호 추출 기법

SOM을 이용하여 데이터를 학습시킨 결과로 얻어지는 출력층 그리드의 승자 뉴런들은 입력 데이터 전체의 패턴을 반영해 줄 수 있는 형태로 대응된다. 이는 SOM 학습 과정인 협동과 적응 과정을 통하여 유사한 속성을 가지는 자료가 모이는 현상으로 발생하는 것이다. 이를 이용하면 주어진 데이터 집합의 특징을 잘 표현하는 일부 데이터를 추출 할 수 있다.

본 연구에서는 정상 주기 신호의 데이터 집단과 이상 주기 신호의 집단을 표현하는 각각의 SOM을 구축한다. 이를 이용하면 각기의 집단을 가장 잘 설명 할 수 있는 특징 데이터가 추출될 수 있다. 전체 데이터를 SOM을 이용하여 학습하게 되면 유사한 신호는 출력층에서 근접한 뉴런으로 묶이게 되고 상대적으로 다른 신호는 멀리 있는 뉴런으로 응집된다.

주기신호의 분석을 위해서 사용되는 특징 신호로는 SOM의 학습과정을 통해 각각의 뉴런에 대항되는 가중치 벡터  $w_j$  를 이용한다. 기존의 과소 표본 추출 방식은 분포를 가장 잘 설명해 줄 수 있는 신호를 데이터 집합에서 선정하는 것이라면 SOM을 이용한 특징 신호 추출 기법은 가중치 벡터를 사용하기에 데이터 전체의 특성을 반영하는 특징 신호를 선정하는 것이라 할 수 있다. 이는 SOM의 학습 과정 중 협동과 적응 과정을 거쳐 승자 뉴런에 대응되는 다수의 데이터를 해당 뉴런의 가중치 값으로 나타낼 수 있기 때문이다. 따라서 뉴런 값을 기반으로 한 특징 신호 추출 방식은 기존 과소 표본 추출 방식보다 데이터의 특징을 표현하는 데 있어서 뛰어나다고 할 수 있다.

정상 신호와 이상 신호의 데이터 비율이 상이한 문제점을 해결하기 위해서 각각의 신호 범주에 대해 동일한 개수의 특징 신호를 추출한다. 이를 위해 두 범주를 표현하는 SOM 출력층의 뉴런 개수를 동일하게 설정하여 동일한 특징 신호를 얻을 수 있도록 하였다. 이상 탐지를 할 때 최대한 많은 정보량을 확보하기 위하여 과소 범주, 주기 신호에서는 이상 신호의 개수인  $n$  개를 기준으로 정사각형으로 그리드 배치를 위해 제공근을 넘지 않는 최대 정수로 한 축의 개수를 설정한다. 이 경우 총  $\lfloor \sqrt{n} \rfloor^2$  개의 뉴런이 설정된다. 예를 들어 과소 범주의 데이터 개수가 30개인 경우 제공근인 5.48보다 작은 최대의 정수인 5가 축의 개수가 되어 총 뉴런이 25개인  $5 \times 5$  그리드 출력층을 설정한다.

그림 2는 본 연구에서 제안하는 특징 신호 추출 과정에 대해 보이고 있다. 이상 신호와 정상 신호의 데이터 집합

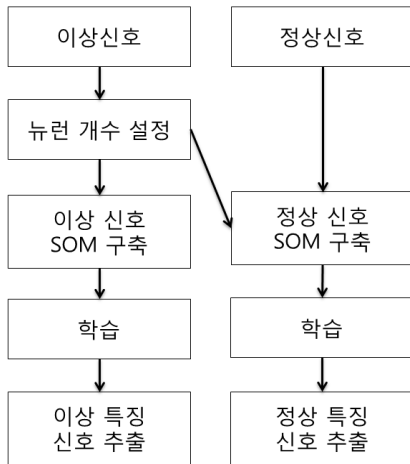


그림 2. 제안 알고리즘 개요

으로부터 특징 신호를 추출하기 위해 이상 신호 데이터의 수를 기반으로 하여 뉴런 개수를 설정한다. 설정된 뉴런 개수를 갖는 각각의 SOM을 구축하고 학습하여 이상 신호와 정상 신호의 특징을 드러내는 신호를 추출하는 과정이다. 알고리즘의 결과로 동일한 개수의 이상 특징 신호와 정상 특징 신호가 얻어진다.

### 3. 이상 탐지

반도체 공정의 생산성을 높이고 제품의 품질을 향상시키기 위해 주기 신호 데이터를 이용한 이상 탐지 및 분류(Fault Detection and Classification)에 관한 연구가 이루어졌다. 주기 신호 데이터의 이상 탐지를 위한 가장 대표적인 기법은 한 시점에서 샘플링 된 자료를 변수로 이용한 Non-stationary SPC 차트이다. 공정이 오랜 시간 동안 계속되는 경우 분석해야 하는 변수가 증가함에 따라 분석 자료의 차원을 줄이기 위한 전처리 기법으로 PCA가 활용되었고 이 외에도 PLS 및 Hotelling  $T^2$  등과 같은 다양한 통계 기법이 이용되고 있다. 통계적 기법뿐만 아니라 기계학습(Machine Learning) 알고리즘, 대표적으로 kNN(k-Nearest Neighbor), 인공신경망 및 SVM(Support Vector Machine) 등을 기반으로 한 이상 탐지 및 분류 연구가 지속적으로 진행 중이다.

본 연구에서는 주기 신호의 이상 탐지를 위해 가장 대표적인 3가지 기법, Hotelling's  $T^2$  관리도, kNN 및 SVM을 이용하여 특징 신호 데이터 집단에 대해 이상 탐지를 진행하고자 한다. Hotelling's  $T^2$  관리도의 경우 다

변량 자료를 한 번에 분석 할 수 있다는 장점이 있고 정상 신호 범주만을 이용하는 가장 대표적인 이상탐지 기법이다(Montgomery, 1996). kNN은 대표적인 예제 기반 학습(instance based learning)기법으로 이상 범주가 존재하는 경우 정상 범주와 비교하여 분석 할 수 있으며 해석이 쉽다는 장점을 지닌다. 또한 SVM은 이상 범주가 존재하는 경우 사용 할 수 있는 가장 강력한 기계학습 알고리즘이며 커널 트릭(kernel trick)의 사용으로 선형 구분자를 비선형으로 확장 할 수 있고 특정 커널을 사용할 경우 인공 신경망과 동일한 효과를 낼 수 있다는 장점을 지닌다(Hastie, et al., 2009).

#### 3.1 Hotelling's $T^2$ 관리도

단변량 SPC차트를 이용하여 공정 주기 신호의 이상탐지를 수행하려면 데이터가 샘플링 되는 각 시점을 하나하나의 변수로 두고 차트를 구성하게 된다. 신호의 길이가 길어지는 경우 관리해야 할 차트의 수가 증가하며 또 각 변수간의 상관관계를 고려하기 어렵다는 점을 지닌다(이재현 등, 2007). 변수간의 상관관계를 고려하여 데이터의 평균에서 떨어진 정도를 하나의 통계량으로 나타내는 가장 대표적인 기법이 Hotelling's  $T^2$  관리도이다.

$$T^2 = (X - \bar{X})^T S^{-1} (X - \bar{X}) \quad (3)$$

식 (3)에서  $\bar{X}$ 와  $S$ 는 기존의 정상 신호 데이터로부터 계산된 표본의 평균과 분산을 나타내며  $X$ 는 이상 여부를 판단하고자 하는 쿼리 데이터(query data)를 의미한다. Hotelling's  $T^2$  통계량은 관측되는 자료가 정상 신호 데이터의 중심점에서 얼마나 떨어졌는지를 각 변수의 상관관계를 고려해서 계산된다. 이때 구한 통계량은 Mahalanobis 거리(Mahalanobis, 1936)의 제곱 값을 가지는데 데이터가 다변량 정규 분포를 따른다는 가정 아래 F분포를 따른다. 정상 데이터의 중심으로부터의 차이가 현격하다면 이상으로 판단해야 하는데 이를 위한 명확한 기준이 필요하다.

$$UCL = \frac{p(m+1)(m-1)}{m^2 - mp} F_{\alpha, p, m-p} \quad (4)$$

식 (4)는 이상을 판단하는 관리한계선(control limit)을 나타내는데 여기서  $p$ 는 변수의 개수,  $m$ 은 총 관측치의 개수를 나타낸다. Hotelling's  $T^2$  통계량은 자유도가  $p, m-p$ 인 F분포를 따르게 된다. 정상 신호 범주에서 발생한 자료임에도 이상라고 분류하는 1종 오류를 범할 확률  $\alpha$ 를 설정할 경우 통계량은 F분포를 따르므로 이상을 판단

하는 기준인 관리한계선을 설정 할 수 있다. 일반적으로 관리한계선은 상한선(upper control limit)과 하한선(lower control limit)이 모두 존재하는데 Hotelling's  $T^2$  통계량이 언제나 양의 값을 갖고 중심으로 부터의 거리가 짧은 것은 문제가 되지 않으므로 따로 설정되지 않는다(강지훈 등, 2011).

Hotelling's  $T^2$  관리도는 여러 개의 변수를 하나의 관리 통계량(monoring statistic)값으로 계산하여 관측 데이터의 이상 여부를 판단한다. 통계량으로 사용되는 값은 단순 거리가 계산되는 것이 아니라 상관관계를 동시에 고려함으로써 데이터의 이상 여부를 판단하는데 있어서 높은 효율성을 가진다.

### 3.2 k-Nearest Neighbor

kNN 알고리즘은 기존 학습 데이터와의 유클리드 거리가 짧은 k개 이웃의 클래스를 바탕으로 다수를 차지하는 클래스로 분류하는 기계학습 알고리즘이다. 판단할 자료와 기준에 보유하고 있는 예제 데이터(instance data)와의 비교를 통해 범주를 구분하므로 예제 기반 학습 기법으로 분류된다. 도출된 결론에 대해 유사 예제를 제공하여 결론을 분석 할 수 있으므로 설명력이 매우 높고 직관적으로 간단히 분류를 수행 할 수 있다는 장점을 지닌다(Hastie, et al., 2009).

$$\hat{Y}(x) = \frac{1}{k} \sum_{x_i \in N_k(x)} y_i \quad (5)$$

$$\hat{C}(x) = \begin{cases} 1 & \text{if } \hat{Y} \geq 0.5 \\ 0 & \text{if } \hat{Y} < 0.5 \end{cases} \quad (6)$$

데이터의 클래스가 0과 1로 나타내 질 때 식 (5)의  $\hat{Y}$ 은 쿼리 데이터  $x$ 와의 유클리드 거리가 짧은 k개의 예제 데이터 집합  $N_k(x)$ 의 클래스 값  $y_i$ 의 평균값을 나타낸다. 식 (6)에 의해  $\hat{Y}$ 값이 0.5보다 크거나 같을 경우 클래스 1로, 0.5보다 작을 경우 클래스 0으로 분류된다. 다시 말해 쿼리 데이터와의 거리가 짧은 k개의 예제 데이터 집합의 클래스 중 가장 빈도가 높은 클래스로 분류하는 것이다.

그림 3에서 k가 5인 경우의 예제를 나타내고 있다. kNN을 이용하면 쿼리 데이터  $x$ 에 대하여 예제 데이터 집합에서 가까운 위치에 있는 5개의 표본을 찾고 빈도가 가장 많이 나타나는 네모 클래스로 분류 할 수 있다. 과거 수집된 정상 및 이상 데이터 집합을 활용하여 새로운 주

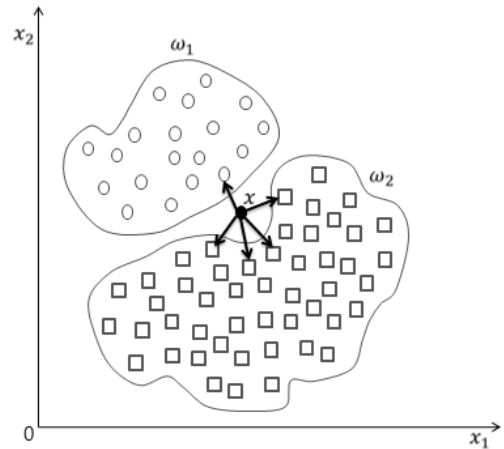


그림 3. kNN 알고리즘 예제

기 신호가 정상인지 이상인지를 판단 할 수 있다.

### 3.3 Support Vector Machine

SVM은 기계학습의 분류 알고리즘 중 하나로써 다른 분류 알고리즘보다 일반화 능력이 뛰어나며 분류 성능이 좋은 것으로 평가된다. 또한 통계적 학습 이론(statistical learning theory)에 기반을 두고 주어진 문제를 항상 전역적 최적해가 보장되는 convex quadratic problem으로 변환하여 해를 구하기 때문에 패턴 인식 분야뿐만 아니라 데이터 마이닝 분야에서 널리 사용되고 있다(Hastie, et al., 2009).

SVM은 학습 데이터가  $\{-1, +1\}$ 처럼 이진 값으로 구분되는 경우, 주어진 데이터의 차원에 존재하는 평면 중 두 데이터 집합으로부터 동일한 거리를 갖으며, 이 거리를 최대화 하는 최대 마진 초평면(maximum-margin hyperplane)을 찾아낸다.

SVM은 클래스에 속해 있는 점들을 분류하는데 있어서 수많은 초평면들 중에 두 클래스의 점들과 거리를 최대한으로 유지하는 최대 마진 초평면을 찾아내는 알고리즘이라 할 수 있다. 이는 아래의 식 (7)에 나타난 최적화 문제를 풀어 해결 할 수 있다.

$$\begin{aligned} \min & \|w\|^2/2 \\ \text{s.t. } & c_i(w \cdot x_i - b) \geq 1, \quad 1 \leq i \leq n \end{aligned} \quad (7)$$

식 (7)에서  $2/\|w\|$ 는 마진을 의미하고 마진을 최대화하는 것은 마진의 역수를 최소화 하는 것과 동일한 의미를 지닌다. 주어진 데이터  $x_i$ 는 -1 또는 1 값을 가지

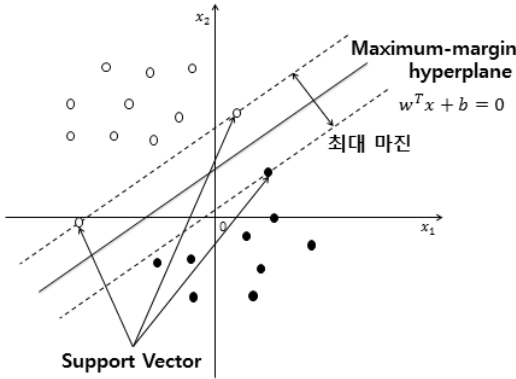


그림 4. SVM의 최대 마진 방식

클래스  $c_i$ 로 구분되고 최대 마진 초평면의 음함수 식값  $w \cdot x_i - b$ 과 같은 부호를 지니도록 설계된다. 이는 최대 마진 초평면을 통해 데이터가 분류 되도록 제약을 설정한 것을 의미한다.

식 (7)의 최적화 문제의 Dual 문제를 구하고 생성되는 입력데이터 간 내적 수식을 커널 함수를 변환한 최적화 문제가 아래의 식 (8)이다(한학용, 2009).

$$\max \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j c_i c_j K(x_i, x_j) \quad (8)$$

$$s.t. \ 0 \leq \alpha_i \leq C, \ i=1, \dots, n$$

$$\sum_{i=1}^n \alpha_i c_i = 0$$

$$f(x) = \text{sgn}(w^T x + b) \quad (9)$$

$$= \text{sgn}\left(\sum_{i=1}^n c_i \alpha_i (x_i^T x) + b\right)$$

새롭게 등장한  $\alpha_i$ 는 마진을 정의한  $w$ 의 Dual 문제 변환 변수라 볼 수 있다. 이 경우 주어진 클래스 판단 연산자는 식 (9)의 첫 번째 식에서 두 번째 식으로 변환다. 식 (7)의 문제에 비해 식 (8)의 문제가 가지는 장점은 데이터가 어떠한 초평면으로 완전히 분리 불가능한 경우에도 가장 성능이 높은 결과를 도출할 수 있으며 정의된 커널함수  $K(x_i, x_j)$ 를 통해 선형으로 존재하는 구분자를 비선형으로 확장할 수 있다는 점이다. 커널함수를 통해 SVM은 비선형 분리 문제를 해결할 수 있고 이를 통해 높은 분류 성능을 지닌다. kNN과 마찬가지로 정상 및 이상 데이터 자료를 이용하여 SVM 이상탐지를 진행할 수 있다.

앞에서 언급 했듯이 주기 신호의 이상 탐지를 하기 위해

대표적인 세 가지 기법인 Hotelling's  $T^2$  관리도, kNN, SVM 알고리즘을 사용 할 것이며 세 가지 기법을 통해서 특징 신호 추출 알고리즘의 성능 우수성을 보이고자 한다.

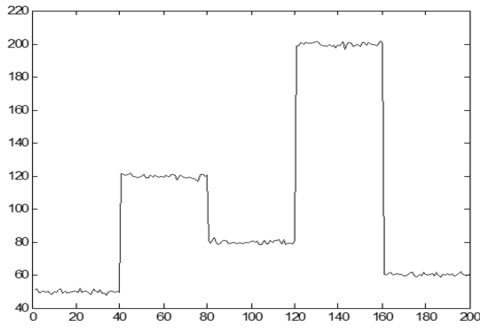
## 4. 실험

특징 신호 추출 알고리즘의 성능을 평가하기 위하여 반도체 공정 신호를 이용한 실험을 진행하였다. 반도체 공정은 크게 웨이퍼 fabrication(이하 FAB 공정), probe 검사(probe test), 조립(assembly), 그리고 최종검사(final test) 등 4단계로 구성되는데(Uzsoy, et al., 1992) 그 중 FAB 공정은 생산되는 반도체의 성능을 좌우하는 가장 중요한 역할을 지닌다. 따라서 FAB 공정의 신호의 이상을 탐지하여 공정 장비를 관리하는 것은 반도체의 품질 향상에 큰 기여를 한다. 보안 문제로 인하여 반출하기 어려운 실제 공정 데이터를 사용하는 대신, 앞선 연구(Park, et al., 2011; 한아향 등, 2010)에서 사용된 반도체 공정 주기 신호 모사 데이터를 이용하였다.

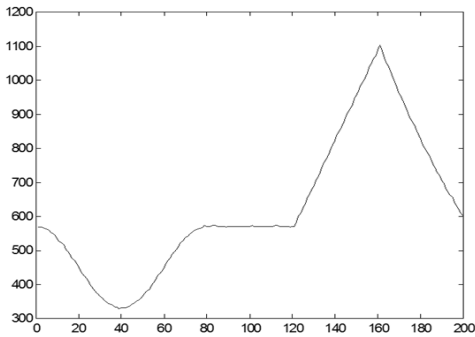
FAB 공정에서 주로 발생하는 세 종류의 신호 형태는 그림 5와 같다. 첫 번째 신호 형태인 step 신호(Type 1)이다. Step 신호는 주로 광전자 센서에서 측정되는 전압과 관련된 데이터이다. 두 번째 신호인 steep 신호(Type 2)는 Diffusion 공정의 온도 센서에서 측정되는 데이터를 모사하였다. 세 번째 신호는 smooth 신호(Type 3)로써 photo-lithography 공정의 온도 데이터이다(한아향 등, 2010). 각각의 신호는 표 1에서 설명한 함수식에 평균이 0이고 표준편차가 1인 정규분포를 따르는 화이트 노이즈(white noise)를 추가하여 실제 공정 데이터에서 발생하는 변동 특성을 반영하였다.

실험 데이터는 실제 공정에서 발생하는 문제인 두 범주 데이터의 비율이 상이한 상황에 맞추어 불량률 5% 정도를 가정하고 1000개의 정상 데이터와 50개의 이상 데이터를 생성하였다. 각 신호의 일정 길이를 각 구간으로 나누어 주었으며 이상 데이터는 공정의 특정구간(4,5구간)에 함수식의 평균 변화를 주어 생성되었다. 평균 변화량은 0.1에서 1까지 0.1씩 증가시켜, 불균형 데이터에서의 작은 변화량부터 큰 변화량까지의 이상 탐지 성능의 민감도를 측정해 볼 수 있도록 하였다.

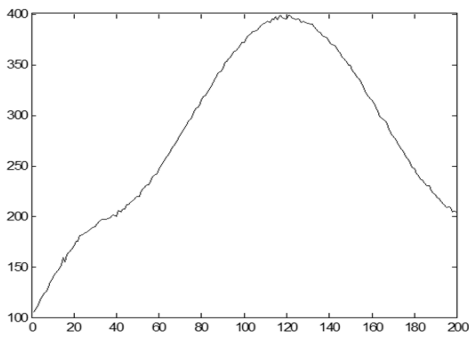
특징 신호 추출 부분에서 설명하였듯, 소수 범주 데이터인 이상 데이터의 개수인 50개에 맞추어  $7 \times 7$  크기를 갖는 SOM 출력층을 설정하고 한 범주로부터 총 49개의 특징 신호를 추출 하였다.



(a) Step 신호



(b) Step 신호



(c) Smooth 신호

그림 5. 주기 신호의 개형

모든 이상 탐지 알고리즘과 마찬가지로 특징 신호 추출 기법을 이용한 이상탐지 알고리즘 또한 1종 오류( $\alpha$ )와 2종 오류( $\beta$ ) 모두를 줄이는 것을 목표로 한다. 다만 불균형 데이터에서의 성능 평가를 위해 일반적으로 데이터의 비율이 비슷한 경우에 성능 평가의 척도로 사용되는 전체 인식률(total accuracy) 대신 정상 데이터의 인식률과 이상 데이터의 인식률을 기반으로 한 기하학적 평균치 계산 방식인 G-Mean(Geometric Mean)을 이용하여 전체 데이터의 정확도를 측정하였다.

일반적으로 전체 데이터에 대한 인식률로 성능을 평가하는 방식을 사용 하지만 데이터의 불균형이 심한 경우에는 이상 데이터의 인식률 보다 정상 데이터의 인식률이 전체 인식률에 차지하는 비중이 크기 때문에 전체 인식률은 각 클래스의 인식률을 제대로 반영 할 수 없다. 따라서 본 연구에서는 성능 평가 방법 중 하나로써 정상 데이터의 인식률  $A_1$ 과 이상 데이터의 인식률  $A_2$ 의 G-Mean값인  $\sqrt{A_1 \times A_2}$ 를 사용하여 각 방법의 성능을 비교한다 (Kang and Cho, 2006).

각각 세 종류의 신호에서 평균 변화량에 따른 9개 이상 데이터 집합을 만들어 정상 데이터 인식률, 이상 데이터 인식률, 기하학적 평균치를 비교하였고 이를 신호의 종류 별로 그림 6, 7, 8에 나타내었다.

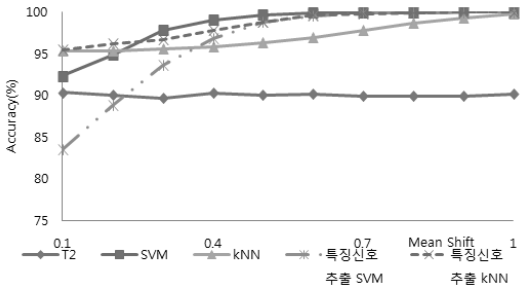
SOM을 이용한 특징 신호 추출 알고리즘의 성능평가를 위해 대조군으로는 Hotelling's  $T^2$  관리도( $\alpha = 0.1$ )를 이용하였다. Hotelling's  $T^2$  관리도는 다변량 이상탐지를 위해 가장 일반적으로 사용되는 방법이며 정상 신호 데이터만을 이용하여 학습하기 때문에 자료가 불균형 데이터인 경우에도 적합하게 작동 할 수 있다. 또한 SVM과 kNN을 특징신호 추출 전 데이터와 추출 후 데이터에 모두 적용하여 특징 신호 추출로 증가하는 인식률에 대하여 살펴보았다.

불균형 데이터의 분류 성능의 척도인 G-Mean 값은 모

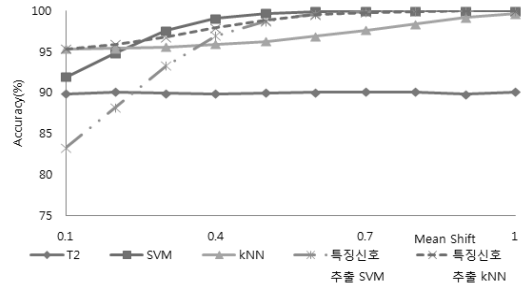
표 1. 실험 데이터 생성 정보

		Type1. Step 신호	Type2. Steep 신호	Type3. Smooth 신호
구간	길이	평균	생성 함수	생성 함수
1	40	50	$450 + 120\cos(x)\pi/40$	$100 + 100\sin(x\pi/80)$
2	40	120	$450 + 120\cos(x)\pi/40$	$97 + 100[\sin(x - 30)\pi/90 + 3\pi/2 + 2]$
3	40	80	570	$97 + 100[\sin(x - 30)\pi/90 + 3\pi/2 + 2]$
4	40	200	$13.25x$	$97 + 100[\sin(x - 30)\pi/90 + 3\pi/2 + 2]$
5	40	60	$200 + 800[\exp - (x - 200)/80]$	$97 + 100[\sin(x - 30)\pi/90 + 3\pi/2 + 2]$

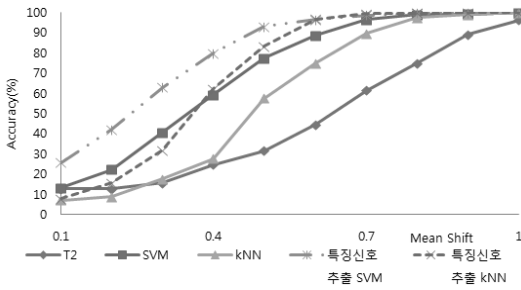




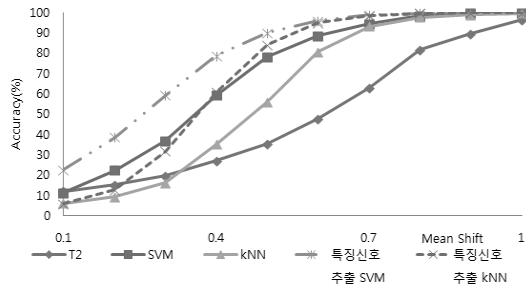
(a) 정상 신호 인식률



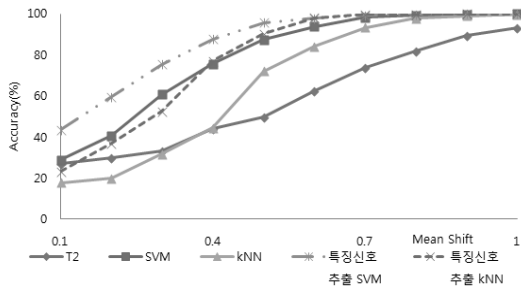
(a) 정상 신호 인식률



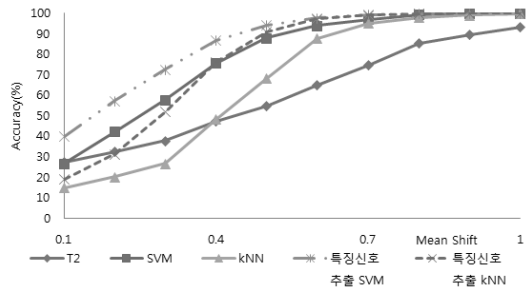
(b) 이상 신호 인식률



(b) 이상 신호 인식률



(c) G-Mean



(c) G-Mean

그림 6. Step 신호에서의 성능 평가 그래프

그림 7. Steep 신호에서의 성능 평가 그래프

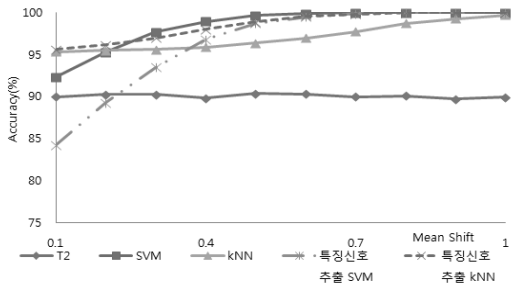
든 신호의 전 구간에서 특징 신호 추출을 거쳐 SVM 학습을 수행한 분류기가 성능이 가장 뛰어남을 확인 할 수 있다. 이는 특징 신호 추출을 통하여 전체 데이터의 특징을 반영하여 각 범주가 가지고 있는 정보량을 동일하게 맞추어 SVM을 일반 자료에 적용한 것보다 높은 성능을 가짐을 확인 할 수 있다. 또 일반적으로 SVM보다 인식률이 떨어지는 kNN 알고리즘의 경우에도 특징 신호 추출을 거쳐 사용한 경우 SVM과 필적하는 성능을 가짐을 확인 할 수 있다.

다만 정상 신호의 인식률의 경우 특징 신호 추출 SVM의 성능이 다소 좋지 못하다. 반면 추출된 특징 신호에 kNN 알고리즘을 적용한 결과가 보다 나은 성능을 보이는 것을 확인 할 수 있는데, SOM을 이용한 특징 신호의 경

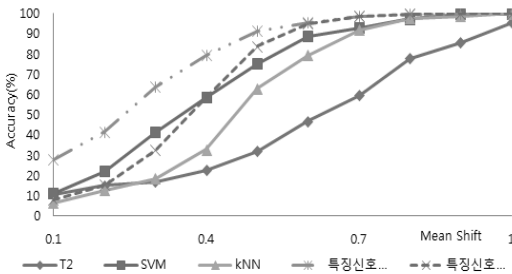
우 데이터 정보가 SOM을 통해 취합되어 나타나는 경향이 있으므로 사례 기반 학습 기법이 보다 나은 성능을 발휘할 수 있다고 판단된다. 일반적인 경우 공정에서의 이상탐지의 가장 큰 이유는 공정의 이상을 조기에 파악하여 불량률의 생산을 방지하는 것이기에 이상 신호의 인식률이 높은 방법이 이상 탐지 기법으로 적합하다고 판단 할 수 있다.

## 5. 결론

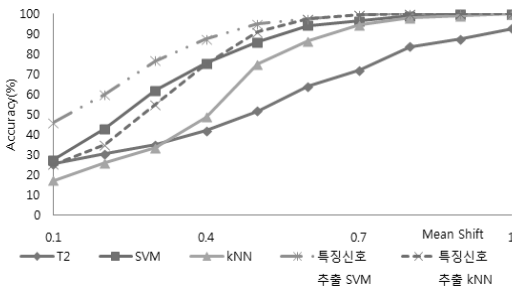
본 연구에서는 실제 공정에서 발생 하는 문제 중에 하나인 정상 데이터와 이상 데이터의 수가 현격히 차이나는 상황에서 이상 탐지가 어려운 문제를 해결하기 위한 방향



(a) 정상 신호 인식률



(b) 이상 신호 인식률



(c) G-Mean

그림 8. Smooth 신호에서의 성능 평가 그래프

을 제시하였다. 정상 데이터와 이상 데이터의 비율을 맞추기 위해 SOM을 구성하여 각각의 데이터의 특징 신호를 추출하였다. 추출된 특징 신호는 기존에 존재하는 데이터 분포의 특징을 잘 나타내는데 있어서 뛰어나다는 장점이 있다. 이를 이용하여 추출된 신호에 SVM과 kNN을 적용한 이상탐지 기법으로 불균형 데이터 상황에서 발생할 수 있는 문제를 해결하였다.

제안한 SOM을 이용한 특징 신호를 추출하여 SVM을 적용하는 경우 정상 데이터의 인식률이 kNN의 인식률 보다 낮은 이유가 데이터의 특징 응집에 따른 현상이라 판단을 하였는데 이 부분에 대한 규명이 필요하다. SOM을 이용하여 추출한 특징 신호의 데이터 분포의 변형에 대한 연구가 진행되어야 정상 신호의 인식률을 증가시키는 것

이 기대된다. 또한 다른 표본 샘플링 기법과의 비교분석을 통해 SOM을 이용한 특징 신호 추출 알고리즘의 다이나믹스 비교 분석은 향후 높은 인식률을 가지는 이상 탐지 알고리즘 개발에 도움이 될 것이다.

## 감사의 글

이 논문은 2011년도 정부(교육과학기술부)의 재원으로 한국연구재단의 기초연구사업 지원을 받아 수행된 것임(2011-0025414). 본 논문은 지식경제부 및 정보통신산업진흥원의 정보통신연구기반구축사업의 연구 결과로 수행되었음(NIPA-2011-(B1110-1101-0002)).

## 참고 문헌

1. 강지훈, 김성범(2011), “붓스트랩을 활용한 이상원인변수의 탐지 기법”, 한국품질경영학회, Vol. 39, No. 2, pp. 234-243.
2. 강필성, 이형주, 조성준(2005), “데이터 불균형 문제에서의 SVM 앙상블 기법의 적용”, 한국정보과학회 가을 학술발표논문집, Vol. 31, No. 2, pp. 706-708.
3. 강필성, 조성준(2006), “데이터 불균형 해결을 위한 Under-Sampling 기반 앙상블 SVMs”, 2006 대한산업공학회/한국경영과학회 춘계 공동학술대회.
4. 박승환, 김준석, 박정술, 김성식, 백준걸(2010), “Support Vector Machine-Regression을 이용한 주기신호의 이상탐지”, 품질경영학회지, Vol. 38, No. 3, pp. 354-362
5. 오장민, 장병탁(2001), “불균형 데이터의 효과적 학습을 위한 커널 퍼셉트론 부스팅 기법”, 한국정보과학회 봄 학술발표논문집, Vol. 28, No. 1, pp. 304-306.
6. 윤재준, 박정술, 백준걸(2010), “SOM을 이용한 주기신호의 이상탐지 및 시각화”, 2010 대한산업공학회 추계 학술대회.
7. 이재현, 김지현, 황지빈, 김성식(2007), “웨이블릿을 이용한 주기 신호 데이터의 이상 탐지에 관한 연구”, 한국시물레이션학회 논문지, Vol. 16, No. 4, pp. 13-22.
8. 한아향, 박정술, 김성식, 백준걸(2010), “시보변 특징점 추출 및 정합을 이용한 주기 신호의 길이 보정 기법”, 한국시물레이션학회 논문지, Vol. 19, No. 4, pp. 111-122.
9. 한학용, 패턴인식 개론, 한빛미디어, 2009.
10. Chawla, N. V., Hall, L. O., Bowyer, K. W., and Kegelmeyer, W. P. (2002), “SMOTE : Synthetic Minority Over-Sampling Technique”, Journal of Artificial Intelligence Research, Vol. 16, pp. 321-357.
11. Chawla, N. A., Lazarevic, A., Hall, L. O., Bowyer, K. W. (2003), “SMOTEBoost : Improving Prediction of the

- Minority Class in Boosting”, Proceeding of the 7<sup>th</sup> European Conference on Principles and Practice of Knowledge Discovery in Databases, Dubrovnik, Croatia, pp. 107-119.
12. Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P. and Uthurusamy, R. (1996), “Advances in Knowledge Discovery and Data Mining”, AAA Press/MIT Press, California.
  13. Hastie, T., Tibshirani, R. and Friedman, J. (2009), “The Elements of Statistical Learning”, Springer.
  14. Japkowicz, N. (2000), “The Class Imbalance Problem : Significance and Strategies”, Proceedings of the 2000 International Conference on Artificial Intelligence, Special Track on Inductive Learning, Las Vegas, Nevada.
  15. Kohonen, T. (1982), “Self-organized formation of topologically correct feature maps”, Biological Cybernetics, Vol. 43, No. 1, pp. 59-69.
  16. Kang, P. S. and Cho, S. J. (2006), “EUS SVMs : Ensemble of under-sampled SVMs for Data Imbalance Problems”, 13th International Conference on Neural Information Processing, Part I, ser. Lecture Notes in Computer Science, Hong Kong, China, Vol. 4232, pp. 837-846.
  17. Laurikkala, J. (2001), “Improving Identification of Difficult Small Classes by Balancing Class Distribution”, Tech Rep. A-2001-2, University of Tampere.
  18. Mahalanobis, P. C. (1936), “On the generalised distance in statistics”, Proceedings of the National Institute of Sciences of India, Vol. 2, No. 1, pp. 49-55.
  19. Montgomery, D. C. (1996), “Introduction to statistical quality control.”, 3<sup>rd</sup> edition, John Wiley and Sons, Inc.
  20. Park, J., Kwon, I. H., Kim, S. S., Baek, J. G. (2011), “Spline regression based feature extraction for semiconductor process fault detection using support vector machine”, Expert Systems with Applications, Vol. 38, No. 5, pp. 5711-5718.
  21. Shi, J. and Jin, J. (2000), “Diagnostic feature extraction from stamping tonnage signals based on design of experiments”, Journal of Manufacturing Science and Engineering, Vol. 122, No. 2, pp. 360-369.
  22. Sun, Y., Kamel, M. S., Wong, A. K. C. and Wang, Y. (2007), “Cost-sensitive boosting for classification of imbalanced data”, Pattern Recognition, Vol. 40, No. 12, pp. 3358-3378.
  23. Uzsoy, R., Lee, C., and Martin-Vega, L. A. (1992), “A Review of Production Planning and Scheduling models in the semiconductor industry PART I: System characteristics, Performance Evaluation and Production Planning”, IIE Transactions, Vol. 24, No. 4, pp. 47-60.



**김 송 이** (kimsongee@korea.ac.kr)

2010 광운대학교 컴퓨터소프트웨어학과 학사  
2010~현재 고려대학교 산업경영공학과 석사과정

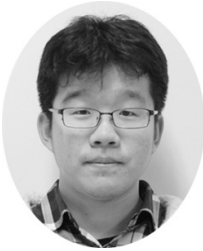
관심분야 : Data mining applications, Applied statistical analysis



**강 지 훈** (joker404@korea.ac.kr)

2009 광운대학교 경영학과 학사  
2011 고려대학교 산업시스템정보공학과 석사  
2011~현재 고려대학교 산업경영공학과 박사과정

관심분야 : Multivariate process control, Data mining algorithms



**박 종 혁** (nrevival@gmail.com)

2007 고려대학교 산업공학과 학사  
2007~현재 고려대학교 정보경영공학대학원 석박사통합과정

관심분야 : Bayesian statistics, Public policies and management



**김 성 식** (sungskim@korea.ac.kr)

1972 고려대학교 기계공학과 학사  
1974 고려대학교 산업공학과 석사  
1976 미국 Southern Methodist University 산업공학과 석사  
1979 미국 Southern Methodist University 산업공학과 박사  
1979~현재 고려대학교 산업경영공학과 교수

관심분야 : Advanced process control, System modeling and optimization



**백 준 걸** (jungeol@korea.ac.kr)

1993 고려대학교 산업공학과 학사  
1995 고려대학교 산업공학과 석사  
2001 고려대학교 산업공학과 박사  
2001~2002 고려대학교 정보통신기술연구소 연구조교수  
2002~2007 인덕대학교 산업시스템경영학과 조교수  
2007~2008 광운대학교 경영학부 조교수  
2008~현재 고려대학교 산업경영공학과 부교수

관심분야 : Advanced process control, Intelligent fault detection, Data mining applications