

## 자기조직화 지도 신경망과 사례기반추론을 이용한 다변량 공정관리

강부식  
목원대학교 경영정보학과  
(bookang@mokwon.ac.kr)

.....

현대의 생산공정에서는 많은 공정변수가 발생하고 있으며 복잡한 연관관계를 가지고 제품의 품질에 영향을 미치고 있다. 따라서 공정의 이상 유무 확인을 위해서는 많은 품질특성치를 동시에 관리하는 다변량 공정관리가 필요하다.

본 연구는 자기조직화 신경망(SOM)과 사례기반추론(CBR) 기법을 이용한 다변량 공정관리 방안을 제안한다. SOM을 이용하여 공정 데이터의 패턴을 생성하고 이상 유무 판단을 위해 기준패턴과 적합성 검정을 한다. 제안한 방법의 검증을 위해 공정에서 발생 가능한 패턴별로 데이터를 생성하여 실험하였고, 실험을 통해 이상패턴을 효과적으로 구별할 수 있음을 보였다. 또한 CBR 방법론을 적용하여 1종 오류는 줄이면서 2종 오류를 아주 작게 유지할 수 있음을 보임으로써, SOM과 CBR 이 결합된 절차가 다변량 공정관리를 위한 한 대안이 될 수 있음을 보였다.

.....

논문접수일 : 2002년 7월

게재확정일 : 2003년 4월

교신저자 : 강부식

### 1. 서론

제조 산업의 경쟁력에 있어 제품의 품질은 가장 기본적인면서도 중요한 역할을 하고 있다. 고객이 요구하는 일정한 품질을 가진 제품을 경제적으로 생산함으로써 품질 원가를 절감할 수 있고, 기업의 경쟁력을 유지하고 강화시킬 수 있다. 생산 현장에서 품질 관리 및 개선을 위해 통계적 공정관리(Statistical Process Control: SPC)가 이루어지고 있다. SPC에서는 공정 흐름에서 발생할 지 모르는 이상 현상을 검출하기 위해 지속적으로 공정 상태를 감시한다. 공정 상태의 감시를 위해 많은 제조 산업의 생산 현장에서 관리도를 사용하고 있다. 관리도는 관리하는 품질특성치의

수에 따라 단변량 관리도와 다변량 관리도로 크게 구분할 수 있다.

실제 생산 현장에서는 생산 기술이 발전하고 제품 구조가 복잡해 짐에 따라 하나의 특성치 보다는 여러 개의 연관된 특성치를 동시에 관리해야 하는 경우가 많아지게 되었다. 반도체와 같은 공정 산업에서는 한 공정에서 수많은 특성치가 발생하고 있고, 제품의 품질은 이런 많은 특성치의 결합에 의해 결정된다. 따라서 효과적인 공정관리를 위해서는 많은 특성치를 동시에 관리하여 공정의 이상 유무를 판단하는 것이 필요하다. 동시에 여러 특성치를 관리하기 위해  $\chi^2$  통계량 혹은 Hotelling의  $T^2$  통계량을 사용하는 다변량 통계적 공정 관리도가 개발되었다 (Kourti and

MacGregor, 1996). 다변량 공정 관리도에는 Hotelling의  $T^2$  관리도(Hotelling, 1947), 다변량 CUSUM(CUmulative SUM) 관리도(Wierda, 1994), 다변량 EWMA(Exponentially-Weighted Moving Average) 관리도(Lowry et al., 1992) 등이 있다. 이러한 기법들은 다변량 정규분포를 기반으로 하여 각 특성치들이 다변량 정규분포를 하고 있으며, 특성치들 간의 상관관계를 알고 있거나 과거 데이터로부터 예측 가능하다고 가정하고 있다(Mason et al., 1997). 그러나 구성하는 특성치의 수가 많은 경우 특성치들 간의 상관관계를 미리 파악하는 것이 힘든 경우나 관리하는 특성치의 수가 증가하게 되면 다변량 정규분포 계산의 복잡도가 크게 증가하게 되어 실무적으로 사용하기에는 많은 어려움이 따르게 된다. 홍창익(1995)은 다변량 관리도를 공정관리를 위해 실제 운용하거나 혹은 이상원인을 해석하는 과정에서 나타날 수 있는 문제점으로 첫째 계산과정이 복잡하고, 둘째 관리도 상에서 관리 한계선을 벗어나지 않는 품질 특성치의 경향이나 주기적인 변동 등을 구별하기 어려우며, 셋째로 평균이나 분산의 작은 변화에 둔감하여 제 2종 오류의 확률이 크다는 점을 들고 있다.

통계적인 기법의 경우에는 다변량의 경우에 변수들의 상관관계를 나타내기 위하여 사전의 정보를 미리 가지고서 변수들의 상관관계를 모델링하여야 하지만 신경망을 이용하는 경우에는 신경망 자체가 변수들의 상관관계까지 학습할 때 고려하기 때문에 사전의 모델링 단계를 거칠 필요 없이 학습에 필요한 변수만을 알아내면 되는 장점이 있다(유성진, 1997). Hwang (1993)은 역전파 신경망을 이용하여 공정 중의 이상패턴을 효과적으로 검출할 수 있음을 보였고, Guh et al. (1999)은 여러 역전파 신경망을 단계적으로 사용

하여 정상 패턴과 이상패턴의 판별, 그리고 이상패턴의 종류를 판별할 수 있음을 실험적으로 제시하였다. Al-Ghanim (1997)은 무감독 자기 조직 신경망인 ART1을 이용하여 관리도의 이상패턴을 식별할 수 있는 방안을 보였고, Pham과 Oztemel (1994)은 관리도의 패턴을 인식하기 위해 LVQ(Learning Vector Quantization) 신경망을 사용할 수 있음과 역전파 신경망을 이용한 기존 연구와 비교하여 효과적으로 패턴을 검출할 수 있음을 보였다. 앞의 여러 연구들은 신경망을 이용하여 공정관리를 효과적으로 할 수 있음을 제시하였으나, 공정관리를 위해 하나의 특성치만을 사용하는 경우에 대한 것으로 관심이 있는 다변량 공정관리에 대해서는 언급하고 있지 않다. 홍창익(1995)은  $T^2$  관리도를 사용하는 경우 관리대상이 되는 품질특성치의 이상원인 해석을 위해 인공신경망을 이용하는 방법을 제안하였다. 유성진(1997)은 다변량 공정관리를 위해 자기 조직 지도 (Self-Organizing Map: SOM) 신경망이 쓰일 수 있음을 보였다. 공정 데이터를 SOM으로 학습하면 학습의 결과 공정 데이터의 특성이 SOM 네트워크의 입력노드와 출력노드를 연결하는 연결가중치로 함축이 되는 특성을 이용하여 다른 공정 데이터를 학습시킨 후 두 연결가중치를 비교하여 두 데이터 셋의 동질성 여부를 파악하고 있다. 동질성 여부의 파악을 위해서는 엔트로피 척도를 도입하고 있다. 휴리스틱한 엔트로피 척도에 의해 두 패턴의 동질성 여부를 0과 1 사이의 값으로 나타내고 있으나 어느 정도의 값 이상이면 동질한 지에 대한 실험 근거는 제시하지 않고 있고, 두 데이터 셋의 패턴 비교를 통해 동질성 여부를 판단하는 방안으로서 SOM의 적용 가능성을 보이고 있다. Kang and Park(2000)은 SOM의 출력패턴을 이용하여 다변량 공정 데

이터의 패턴을 판별할 수 있음을 두 데이터 셋의 실험을 통해 제시하였다. 그러나 공정 중에 발생 가능한 상승(하강) 이동 추세, 상향(하향) 추세, 주기적 추세에 대한 검증은 하고 있지 못하다.

이 연구에서는 Kang and Park(2000)의 연구를 바탕으로 공정 중에 발생 가능한 다양한 패턴을 SOM을 이용하여 검출할 수 있음을 보이고자 한다. 또한 다변량 공정관리를 위해 SOM을 사용시에 현장에서 적용 가능한 절차를 사례기반추론 방법론에 따라 제시하고자 한다. 이 연구는 전체 6장으로 구성되어 있다. 2장에서는 연구에서 사용하는 SOM, 적합성 검정, 사례기반추론에 대해 살펴보고, 3장에서는 다변량 공정관리를 위한 SOM과 적합성 검정 기법을 사용한 패턴 검출 방법론에 대해 설명한다. 4장에서는 공정 중에 발생 가능한 다양한 패턴 데이터 셋을 생성하여, 제안한 방법론에 따라 실험하고 그 결과를 평가한다. 5장에서는 사례기반추론 방법론에 따른 다변량 공정관리 절차를 소개하고, 6장에서 결론을 맺고 향후 연구사항을 제시한다.

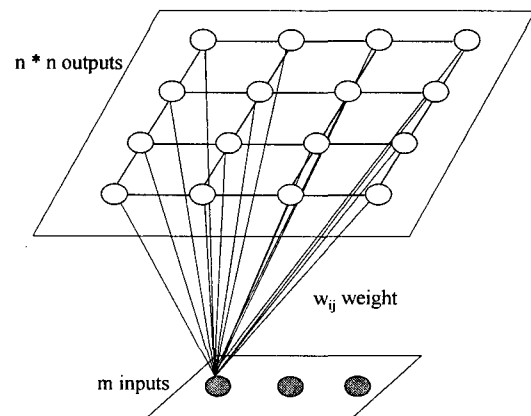
## 2. 관련 기법

### 2.1 SOM(Self-Organizing Map) 신경망

SOM 신경망은 Kohonen(1990)이 제안한 무감독 학습기법으로, <그림 1>과 같이 구성된다. SOM은 입력층과 출력층으로 구성된 신경망으로, 입력층의 입력노드수는 입력벡터 차원의 수  $m$ 으로, 출력층의 노드 수는  $N \times N$ 의 격자모양의 개수를 지정한다. 학습은 입력층과 출력층을 연결하는 연결가중치의 값을 조정하는 것이다.

학습이 끝나면 비슷한 성격의 입력벡터 값을

가지는 데이터들은 출력층의 특정노드에 맵핑이 된다. 따라서 비슷한 성격의 입력 데이터가 자연스럽게 클러스터링이 된다. 만약에 두 데이터 셋이 유사한 입력벡터로 구성되어 있다면 학습된 SOM에 대한 출력노드에서의 맵핑 형태는 유사한 분포를 갖게 될 것이다. 즉 데이터 셋간의 유사성을 SOM의 출력노드의 분포(이를 패턴이라 부르기로 한다)를 통해 결정할 수 있게 된다. 또한 SOM의 특성중 하나는 SOM을 학습시에 학습용 데이터간의 관계에 대한 사전지식이 없어도 SOM의 학습 규칙에 따라 연결가중치가 학습이 되기 때문에 복잡한 연관관계가 있는 데이터 셋에 대해 쉽게 학습이 가능하다는 점이다.



<그림 1> SOM 신경망의 구성 형태

SOM의 학습 알고리즘은 내부에서 사용하는 유사성의 척도 및 이웃함수, 그리고 학습 파라미터의 결정 등에 따라 약간의 차이가 있으나 다음과 같이 이루어진다 (Bigus and Bigus, 2001). 먼저, 입력 데이터 셋이 입력층에 전달된다. 다음에, 입력벡터  $x$ 와 출력노드  $j$ 를 연결하고 있는 연결가중치 벡터  $w_j$ 와의 거리를 유클리디안 거리

로 계산한다.

$$y_j = \|x - w_j\|^2 \quad (1)$$

가장 적은  $y_j$ 을 가진 출력노드  $j$ 를 승리 노드로 선정한다. 승리 노드와 승리 노드의 이웃 노드의 연결가중치를 다음 식을 따라 조정한다.

$$w_j(t+1) = w_j(t) + \beta(k)C_{ij}(k)y_j(t) \quad (2)$$

위 식에서  $\beta(k)$ 는 학습 반복  $k$  단계에서의 학습율을,  $C_{ij}(k)$ 는 학습 반복  $k$ 단계에서 노드  $i$ 와  $j$ 의 이웃 함수 값을,  $y_j$ 는  $t$  시점에서 입력벡터  $x$ 와 가중치 벡터  $w_j$ 사이의 유클리디안 거리를 나타낸다.  $C_{ij}(k)$ 는 Gaussian 함수로 멕시코 모자 형태를 취하고 있는데, 다음과 같이 정의된다.

$$C_{ij}(k) = \exp\left(-\frac{\|i - j\|^2}{\sigma(k)^2}\right) \quad (3)$$

위 식에서  $i, j$ 는 SOM의  $N \times N$  2차원 출력노드 맵에서의 좌표, 즉 출력노드  $i$ 와  $j$ 의 좌표를 나타내고,  $k$ 는 학습 단계를 나타낸다. 이웃 함수의 넓이는  $\sigma(k)^2$ 으로 학습 단계  $k$ 가 적을 때는 출력노드 전체에 해당할 만큼 넓게,  $k$ 가 최대값에 이르면 출력노드 중 한 노드만 해당이 되도록 정의한다. 즉 학습이 진행될수록 점차 이웃함수의 범위를 좁히게 된다.

$\beta(k)$ 는 다음과 같이 계산된다.

$$\beta(k) = \beta_{initial} \left( \frac{\beta_{final}}{\beta_{initial}} \right)^{\frac{k}{k_{max}}} \quad (4)$$

$k_{max}$ 는 수행해야 할 최대 학습 반복회수를 의미한다. 학습율  $\beta(k)$ 는  $k$ 가 증가함에 따라 지수

적으로 감소한다. 일반적으로  $\beta_{initial}=1.0$ 을,  $\beta_{final}=0.05$ 를 사용한다.

## 2.2 적합성 검정

적합성 검정은 기대도수와 관측도수 간의 데이터 분포를 비교하여 관측도수와 기대도수와 동일성 여부에 대한 검정을 한다. 기준이 되는 패턴을 구한 다음 현재의 패턴을 기준패턴과 적합성 검정을 하게 되면 기준 패턴과의 동일성 여부를 판정할 수 있다. <표 1>은 두 데이터 셋의 범주별 데이터 분포를 보여 주고 있다. 한 데이터 셋 내의 각 인스턴스는 여러 범주 중에서 하나의 범주에만 속하게 된다. 도수  $O_i$ 는 현재 데이터 셋에서 범주  $G_i$ 에 속하는 인스턴스의 수이다. 현 데이터 셋의 전체 인스턴수의 수가  $n$ 일때,  $O_1+O_2+\dots+O_r=n$ 이 된다.

<표 1> 범주별 데이터 분포

	범주					합계
	$G_1$	$G_2$	$G_3$	...	$G_r$	
관측치(공정패턴)	$O_1$	$O_2$	$O_3$	...	$O_r$	$n$
기대치(기준패턴)	$E_1$	$E_2$	$E_3$	...	$E_r$	$n$

현재 공정패턴(관측 도수)의 한 인스턴스가 범주  $G_i$ 에 속할 확률이  $P_{i0}$  라면,  $P_{10}+P_{20}+\dots+P_{r0}=1$ 이다. 이 경우 귀무가설과 대립가설은 다음과 같다.

$$H_0: (P_1, P_2, \dots, P_r) = (P_{10}, P_{20}, \dots, P_{r0})$$

$$H_1: (P_1, P_2, \dots, P_r) \neq (P_{10}, P_{20}, \dots, P_{r0})$$

각 범주  $i$ 에 속하는 기대도수는  $E_i = n \cdot P_i$  가

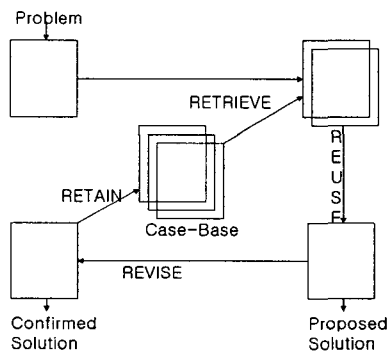
되고, 관측도수는  $O_i = n \cdot P_{i0}$  가 된다. 이때 검정 통계량은 다음과 같다.

$$\chi^2 = \sum_{i=1}^r \frac{(O_i - E_i)^2}{E_i} \quad (5)$$

검정 통계량의 자유도는  $(r-1)$ 이고, 귀무가설의 기각역은 유의수준  $\alpha$ 에서  $\chi^2 \geq \chi^2_{(r-1), \alpha}$  이다.

### 2.3 사례기반추론

사례기반추론(Case-Based Reasoning: CBR)은 과거 문제의 해를 사용하거나 적용시켜서 새로운 문제를 해결하기 위한 방법론(Watson, 1999)으로, 다음 <그림 2>처럼 쉽고 간단한 4가지 활동으로 구성되어 있다(Aamodt and Plaza, 1994).



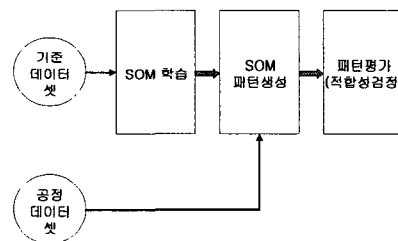
<그림 2> 사례기반추론 방법론

먼저, 새로운 문제에 가장 유사한 사례를 추출하고, 문제를 해결하기 위해 추출된 사례의 해를 사용한다. 필요하다면 추출된 해를 새로운 문제에 맞게 개정한다. 새로운 사례인 경우 사례베이스

에 저장한다. CBR은 하나의 방법론으로서 다양한 기계학습 기법을 4가지 활동 각각에 적합하도록 적용하는 것이 가능하다(Watson, 1999). 이 연구에서는 CBR을 이용하여 다변량 공정관리를 하기 위한 절차를 제안한다. CBR을 이용하여 초기에 확인 가능한 패턴을 기준 사례로 하여 공정관리를 진행하고, 새로운 패턴이 발생하는 경우 사례베이스에 새로운 사례로 저장하고 공정관리를 계속 진행한다.

### 3. SOM과 적합성 검정을 이용한 다변량 공정관리

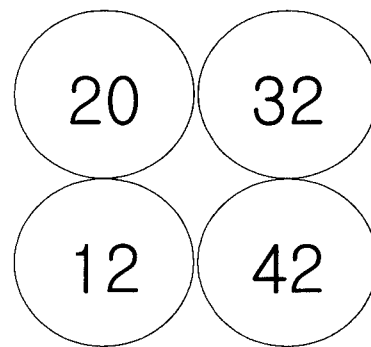
SOM은 2장에서 설명한 것처럼 입력 데이터 셋을 입력으로 받아, 격자형의 출력 형태( $N \times N$ )로 표현할 수 있다. 두 입력 데이터 셋이 다른 경우 나타나는 출력 형태는 달라지게 된다. 이때 적합성 검정을 이용하면, 두 데이터 셋에 대한 SOM의 출력패턴을 비교하여 동일성 여부를 판단할 수 있다. 따라서 3절에서는 SOM과 적합성 검정을 사용하여 공정 데이터의 이상 유무를 판단하기 위한 방법을 제시하고자 한다 (<그림 3> 참조).



<그림 3> SOM과 적합성 검정을 이용한 데이터 셋의 유사성 판단

기준 데이터 셋은 이미 데이터 셋의 특성을 알고 있는 데이터 군이다. SOM의 학습 단계에서는 기준 데이터 셋을 이용하여 SOM을 학습시킨다. SOM 신경망에서 입력층의 입력노드는 기준 데이터 셋의 차원(변수)의 수로, 출력층의 출력노드 수는 격자형의  $N \times N$ 의 형태로 구성한다. 이때  $N$ 은 적용되는 응용 분야에 따라 결정한다. SOM의 구조를 결정한 후 2절에서 설명한 알고리즘에 따라 SOM을 학습시킨다. 학습이 끝나면 입력노드와 출력노드를 연결하는 가중치 값이 결정이 된다. 패턴 생성 단계에서는 학습된 SOM을 이용하여 출력노드의 패턴을 생성한다. 하나의 입력벡터  $x_i$ 를 학습된 SOM에 입력하면 SOM의 연결가중치  $w_{ij}$ 와의 유클리디안 거리가 계산이 되어지고 그 결과 출력노드  $j$ 의 출력값이 결정된다. 출력값을 1-유클리디안 거리로 정하면, 출력노드  $j$ 의 값이 클수록 입력벡터  $x_i$ 는 출력노드  $j$ 에 그만큼 강한 관계가 있다는 것을 나타낸다. 일반적으로는 가장 큰 출력값을 갖는 출력노드에 입력벡터가 속하는 것으로 하여 클러스터링 도구로 SOM을 많이 사용하지만, 이 연구에서는 입력벡터에 대해 일정한 출력 값  $C$  이상을 갖는 출력노드 모두에 입력벡터가 속하는 것으로 정의한다. 왜냐하면 공정의 변화에 따라 공정 데이터는 영향을 받게 되고, SOM의 각 출력노드의 출력값에 약하게 혹은 강하게 영향을 주게 되는 데 상대적인 비교에 의한 한 출력노드의 선정은 이런 공정 데이터의 변화를 잘 검출하지 못하기 때문이다. 이는 이 연구의 실험과정에서 확인을 하였다. 패턴 생성 단계에서는 입력 데이터 셋의 각 입력벡터(인스턴스)에 대해  $C$  이상의 출력값을 갖는 출력노드에 각 인스턴스가 속하는 것으로 하고, 각 출력노드에 속한 인스턴스의 분포를 패턴이라 명명한다. 다음 <그림 4>는 출력노드가  $2 \times 2$ 인 패

턴의 한 예이다. 첫번째 출력노드의 20은 입력 데이터 셋의 인스턴스 중에서 첫번째 출력노드의 출력값이 기준값  $C$  이상인 인스턴스가 20개임을 나타낸다.



<그림 4> 출력패턴의 예

패턴 평가 단계에서는 기준 데이터 셋의 기준 패턴과 공정 데이터 셋의 공정패턴을 적합성 검정을 통해 비교한다. 적합성 검정은 각 출력노드 별로 판정한다. 그러기 위해서 원래 출력노드의 패턴을 <표 2>와 같이 이항 형태로 변환한다. 만약 기준과 입력 데이터 셋의 인스턴스 개수가 50이고 SOM의 출력노드에 대한 기준패턴이 15, 30, 14, 40이라면, <그림 4>의 첫번째 출력노드에 대한 적합성 검정을 위한  $\chi^2$  통계량은  $(20-15)^2 / 20 + ((50-20)-(50-15))^2 / (50-15) = 2.381$ 로, 자유도가 1이고 유의수준 0.05인  $\chi^2$  값은 3.8415보다 적다. 따라서 유의수준 0.05로 공정 데이터의 첫번째 출력노드의 분포는 기준 데이터의 첫번째 출력노드의 분포와 다르지 않다고 판단한다. 마찬가지로 나머지 출력노드에 대해서도 적합성 검정을 한다. 적합성 검정의 결과 모든 출력노드의 분포가 하나라도 기각되지 않는 경우 현재 패턴과 기준패턴은 동일하다고 판정을 한다.

<표 2> 출력노드별 이항 형태로의 변환  
(데이터 셋의 크기가 50인 경우)

(a) 원래 패턴

	출력노드			
	(0,0):A	(0,1):B	(1,0):C	(1,1):D
공정패턴	20	32	12	42
기준패턴	15	30	14	40

(b) 출력노드별 이항 형태로의 변환

A	~ A	B	~ B	C	~ C	D	~ D
20	30	32	18	12	38	42	8
15	35	30	20	14	36	40	10

#### 4. 실험 및 결과 분석

SOM을 이용한 다변량 공정관리의 가능성을 실험하기 위해 공정에서 자주 발생 가능한 여섯 종류의 패턴에 대해 MS-Excel의 난수 생성 기법을 이용하여 실험 데이터를 인위적으로 생성하였다. 각 패턴의 생성은 Guh et al. (1999)이 제안한 식을 참조하였다.

정규패턴:  $x_i(t) = n_i(t)$ ,  $n_i(t) \sim N(0,1)$  (6)

상향 혹은 하향 이동 패턴:  $x_i(t) = n_i(t) + u \times s$  (7)

증가 혹은 감소하는 추세 패턴:

$$x_i(t) = n_i(t) + d \times t \quad (8)$$

주기 패턴:  $x_i(t) = n_i(t) + a \times \sin(2\pi t / \Omega)$  (9)

이 때 t는 시점을 나타내며  $1 \leq t \leq 50$ 의 범위를 갖는다.  $x_i(t)$ 는 시점 t의 변수  $x_i$ ,  $i=1, \dots, m$ 를 위한 샘플 값이고,  $n_i(t)$ 는 시점 t와 변수  $x_i$ ,  $i=1, \dots, m$ 인 경우의 평균이 0이고 분산이 1인 정규분포를, u는 이동의 위치를 결정하기 위한 파라미터를 (u=0은 이동이 없음을 u=1은 이동이 있음을 의미), s는 이동의 크기를 ( $-2 \leq s \leq 2$ ), d는 추세 기울기 ( $-0.08 \leq d \leq 0.08$ ), a는 주기의 진폭을 ( $1 \leq a \leq 2$ ) 그리고  $\Omega$ 는 주기의 기울기 ( $10 \leq \Omega \leq 50$ ) 나타낸다.

한 데이터 셋은 6개의 변수를 가진 50개의 인스턴스로 구성하였다. 한 데이터 셋의 예는 <표 3>과 같다.

6개 패턴별로 데이터 셋을 15개씩 생성하여 총 90개 데이터 셋을 준비하였다. 예를 들어 정규패턴의 경우 각 변수별로 seed를 달리하여 평균이 0이고 분산이 1인 정규분포에서 750개씩 난수생성을 한 후, 50개 단위로 묶으면 15개의 데이터 셋을 구성할 수 있다. 각 이상패턴은 또다른 seed를 사용하여 난수생성한 정규분포에서 적절한 s,d,a,(를 가지고 생성한다. SOM을 이용한 패턴 검출은 기준패턴을 먼저 생성하고 나머지 패턴들과 기준패턴을 비교함으로써 성능을 알아보도록

<표 3> 데이터 셋의 구성 예

	var1	var2	var3	var4	var5	var6
instance1	-3.02301	-2.99478	-2.97499	-2.95637	-2.93876	-2.91649
instance2	0.160065	1.235237	-0.77351	0.119737	1.153599	-0.82865
instance3	-0.86578	0.638688	-0.57207	0.94964	-0.32127	1.399626
...	...	...	...	...	...	...
instance50	-0.80781	1.27251	0.220286	-0.59492	1.808698	0.39485

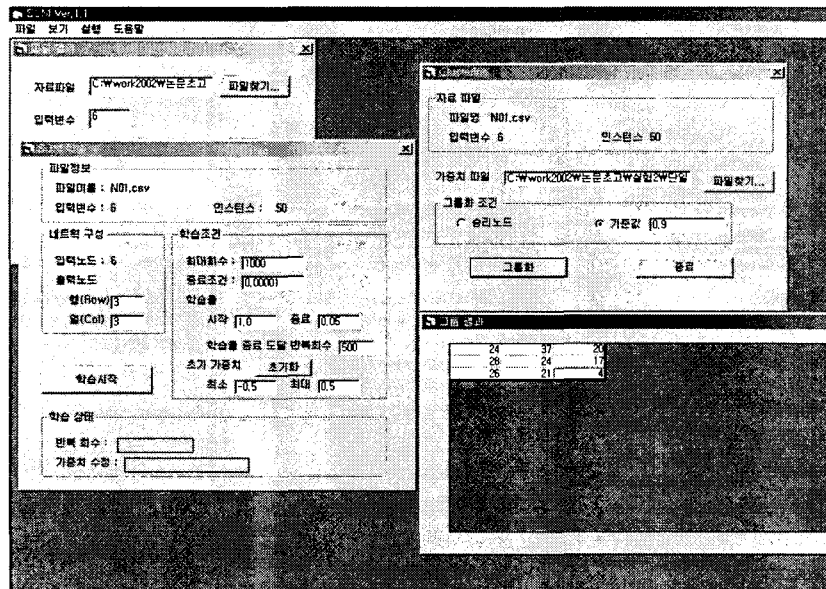
한다. 기준패턴은 각 패턴의 데이터 셋 중에서 하나를 선택하여 기준패턴을 생성하고 나머지 다른 패턴을 비교하도록 한다.

실험을 위하여 먼저 SOM 알고리즘을 Bigus and Bigus(2001)가 제시한 자바 프로그램에 필요한 기능을 추가하여 비주얼 베이직으로 작성하였다.

<그림 5>는 SOM 학습 조건 화면, SOM의 패턴 생성 조건 화면 및 패턴 생성 결과를 보여주고 있다. SOM의 N\*N 출력노드 구성에서 N의 결정은 응용 영역에 따라 조금씩 달라지게 된다. 응용 분야의 데이터 특성에 따라 좋은 특성을 보이는 N이 다를 수 있기 때문이다. 이 실험에서는 고려하는 패턴의 수가 6개임으로 패턴의 수보다는 많게 N=3으로 하였다. 그리고 출력노드에 속하는 것으로 판정하는 출력의 기준값에 따라서 출력노드의 분포가 영향을 받게 되는 데, 데이터 셋의 전체 인스턴스 중에서 20~80%에 해당하는

인스턴스가 각 출력노드에 속하도록 분포되는 기준값 0.9를 설정하였다. 기준값이 너무 낮으면 거의 모든 인스턴스가 각 출력노드에 속하게 되고, 따라서 공정 데이터의 이상 현상이 내포시 출력노드의 분포에 적게 영향을 미치지 때문이다. 마찬가지로 너무 높은 기준값도 제대로 공정 데이터의 특성을 반영하지 못하기 때문이다.

패턴의 동일성 여부 판정을 위한 적합성 검정시 응용 분야의 데이터 특성에 따라 적당한 유의수준의 선정이 필요한데 이 실험에서는 유의수준 0.01과 0.05에 대해 살펴본다. 다음 <표 4>와 <표 5>는 정규패턴과 이상패턴을 기준패턴으로 한 경우 SOM과 적합성 검정을 이용한 동일성 여부 판정에 대한 실험결과이다. <표 4>는 정규패턴을 기본 패턴으로 한 경우에 정규패턴과 다른 패턴을 유의수준 0.01과 0.05에서 적합성 검정한 결과이다. <표 5>는 <표 4>의 각 이상패턴을 기준패턴으로 한 경우, 기준패턴인 이



<그림 5> SOM 프로그램



<표 4> 기준패턴이 정규패턴인 경우 적합성 검정결과 (기준값=0.9일때)

패턴	데이터셋	유의수준 0.05		유의수준 0.01	
		Accept	Reject	Accept	Reject
정규패턴	15	3	12	11	4
상향이동패턴 (u=1, s=1)	15	0	15	0	15
하향이동패턴 (s=1, s=-1)	15	0	15	0	15
증가추세패턴 (d=0.05)	15	0	15	0	14
감소추세패턴 (d=-0.05)	15	0	15	0	15
주기패턴 (a=1, Q=25)	15	0	15	1	14
합	90	3	87	12	78
	1종 오류	0.8		0.27	
	2종 오류	0		0.013	

<표 5> 기준패턴을 이상패턴으로 한 경우 적합성 검정결과 (기준값=0.9, \*표시가 기준패턴)

패턴	데이터셋	유의수준 0.05				유의수준 0.01			
		Accept	Reject	1종오류	2종오류	Accept	Reject	1종오류	2종오류
*상향이동패턴	15	8	7	0.47	0	12	3	0.2	0
그외 패턴	75	0	75			0	75		
*하향이동패턴	15	9	6	0.4	0	13	2	0.133	0
그외 패턴	75	0	75			0	75		
*증가추세패턴	15	8	7	0.47	0	13	2	0.133	0
그외 패턴	75	0	75			0	75		
*감소추세패턴	15	7	8	0.533	0	14	1	0.07	0
그외 패턴	75	0	75			0	75		
*주기패턴	15	7	8	0.533	0	13	2	0.133	0.093
그외 패턴	75	0	75			7	68		

상패턴과 다른 패턴과의 적합성 검정에 대한 결과이다.

<표 4>와 <표 5>에서 유의수준 0.01에서 1종 오류가 감소하는 반면 2종 오류가 증가하고, 유의수준 0.05에서는 1종 오류가 크게 증가하나 2종 오류는 0으로 나타났다. 따라서 1종 오류와 2종 오류를 고려하여 적절한 유의수준의 선정이 필요하다라는 것을 실험을 통해 알 수 있다. 위 실험 결과에서 유의수준이 0.05일 때, SOM의 출력패

턴은 입력 데이터 셋의 변화에 대해 상당히 민감하게 변함을 알 수 있다. 공정변화를 민감하게 잘 검출하는 반면 1종 오류가 크게 나타나는 문제가 있다. 이 경우 1종 오류를 줄일 수 있는 방안이 모색이 된다면, 동시에 고려해야 할 변수의 수가 많고 실시간으로 공정 데이터 셋에 대한 정상 혹은 이상 여부를 판정하고자 하는 다변량 공정관리에서 효과적인 대안이 될 수 있음을 위 실험에서 보여 주고 있다. 1종 오류를 줄이는 방법의 하

나로 CBR 방법이 사용될 수 있다. 위 <표 4>의 정규패턴 경우에서 1종 오류가 발생된 정규패턴이 이상패턴으로 판정되지 않고 또 다른 하나의 정규패턴으로 판정되어 정규패턴의 기준패턴으로 추가될 수 있다면 점차 1종 오류를 감소시킬 수 있을 것이다. CBR을 이용한 절차에 대해서는 5절에서 좀 더 상세하게 설명이 이루어지는데, 간략하게 단계를 기술하면 다음과 같다.

- 1) 정규 기준패턴과 공정패턴을 적합성 검정한다. 적합성 검정 결과 기각되면 다음 단계로 진행한다. 기각되지 않으면 공정패턴을 정규패턴으로 판정한다.
- 2) 공정패턴을 이상 기준패턴과 적합성 검정한다. 적합성 검정 결과 기각되면 3단계로 진행한다. 기각되지 않으면 유사 이상 기준패턴으로 공정패턴을 판정한다.
- 3) 현 공정패턴은 사례베이스내의 기준패턴 어느

것과도 적합하지 않으므로 현장 엔지니어의 분석 단계로 넘어가서 정규패턴 혹은 이상패턴으로 판정을 받는다.

- 4) 새로운 정규패턴 혹은 이상패턴을 사례베이스에 추가한다.

<표 4>의 정규패턴 데이터 셋 15개에 대하여 위에 제시한 4단계의 절차에 의해 적용해 본 결과는 <표 6>과 같다.

<표 6>의 적용실험에서 추가로 살펴봐야 하는 것은 새로운 정규 기준패턴으로 각 공정패턴이 사례베이스에 추가될 경우에 <표 4>의 각 이상패턴이 새로이 추가된 정규 기준패턴과의 적합성 검정에서 기각되는지의 여부이다. 새로운 정규 기준패턴이 추가된 경우의 이상패턴에 대한 2종 오류는 실험결과 0으로 나타났으며, 정규패턴의 새로운 기준패턴으로의 추가가 전체 패턴에 대해 1종 오류는 줄이면서 2종 오류는 0으로 유

<표 6> 정규패턴 데이터 셋의 5단계 절차 적용결과

정규 데이터 셋	단계			
	1	2	3	4
N01	정규패턴 판정			
N02	정규패턴 기각	이상패턴 기각	정규패턴 판정	기준패턴 추가 (2종 오류=0)
N03	정규패턴 판정			
N04	정규패턴 기각	이상패턴 기각	정규패턴 판정	기준패턴 추가 (2종 오류=0)
N05	정규패턴 기각	이상패턴 기각	정규패턴 판정	기준패턴 추가 (2종 오류=0)
N06	정규패턴 기각	이상패턴 기각	정규패턴 판정	기준패턴 추가 (2종 오류=0)
N07	정규패턴 판정			
N08	정규패턴 판정			
N09	정규패턴 기각	이상패턴 기각	정규패턴 판정	기준패턴 추가 (2종 오류=0)
N10	정규패턴 판정			
N11	정규패턴 판정			
N12	정규패턴 판정			
N13	정규패턴 판정			
N14	정규패턴 기각	이상패턴 기각	정규패턴 판정	기준패턴 추가 (2종 오류=0)
N15	정규패턴 판정			

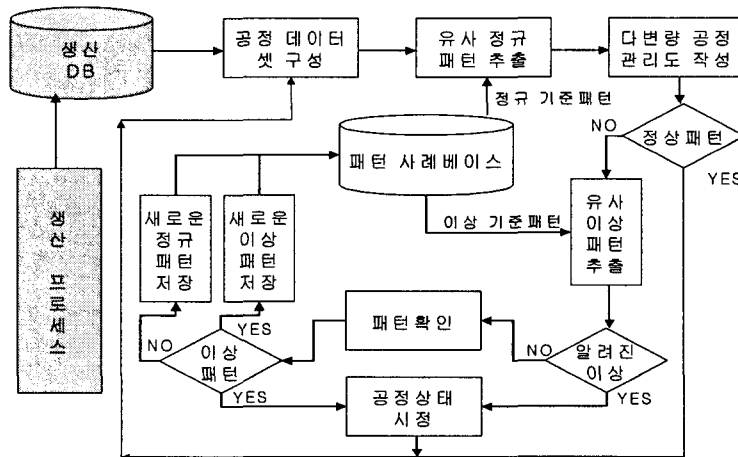
지하는 것을 알 수 있었다.

실험 결과 SOM과 사례기반추론을 결합한 기법이 다변량 공정관리에 적용 가능한 한 대안이 될 수 있음을 알 수 있다. 다만 이상패턴의 경우  $s, d, a, \Omega$ 와 같은 파라미터의 변화에 따라 다양한 형태의 패턴이 존재하기 때문에 충분한 실험과 분석을 통해 이상패턴에 대한 사전 정의가 필요하다.

### 5. 사례기반추론을 이용한 다변량 공정관리 방안

이 절에서는 3장에서 설명한 SOM과 적합성 검정을 이용하여 패턴을 구별하는 방법을 이용하여 어떻게 다변량 공정관리를 할 것인지에 대한 방안을 사례기반추론 방법론의 골격을 따라 <그림 6>처럼 제안한다. 생산 프로세스에서는 공정 제품에 대한 특성치가 지속적으로 생산 데이터베이스에 저장된다. 다변량 공정관리에서 공정 상태를 확인하기 위해 특성치를 실시간으

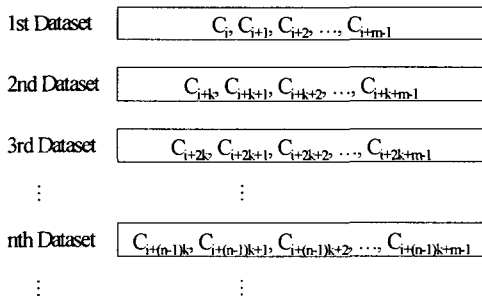
로 구성하여 알려진 패턴과 비교하게 된다. 패턴 사례베이스에는 정상패턴과 이상패턴에 대해 학습된 SOM의 가중치 파일 정보와 출력패턴 정보가 저장되어 있다. 사례베이스내 유사 기준패턴을 추출하는 과정은 2단계로 이루어진다. 먼저 정규패턴에 대해 유사 패턴을 추출하고 적합성 검정 결과 정규패턴이 아닌 경우 유사 이상패턴을 추출하여 적합성 검정을 하도록 한다. 이는 공정상의 패턴을 정규패턴과 이상패턴으로 단순화하여 분류한 후에 이상패턴인 경우 다음 단계에서 이상패턴의 타입을 검출하는 것이 효과적이라는 Guh et al.(1999)의 연구 결과를 반영함과 동시에 이상패턴의 경우 파라미터의 값에 따라 매우 다양한 패턴이 존재하기 때문에 정규 및 이상패턴을 동시에 고려하여 공정 데이터 셋에 가장 유사한 기준패턴을 추출하게 되면 유사 패턴 추출시의 계산의 효율성이 저하되는 것을 방지하기 위해서다. 사례기반추론 방법론에 따른 다변량 공정관리 흐름도의 주요 단계에 대한 설명은 다음과 같다.



<그림 6> 사례기반추론 방법론에 따른 다변량 공정관리 흐름도

### 5.1 공정 데이터 셋 구성

생산 공정에서 수집된 공정 특성치를 추출하여 공정 데이터 셋을 구성한다. 기준패턴을 생성시의 데이터 셋과 동일하게 공정 데이터 셋을 구성한다. <그림 7>은 연속 공정에서 공정 데이터 셋을 구성하는 예를 보여주고 있다. 연속 공정에서의 공정 상태를 확인해야 하기 때문에 이전의 데이터 셋에 새로운 k개 인스턴스를 교체하여 공정의 이상 유무를 확인한다. k의 값이 작으면 공정의 이상 여부를 자세하게 관찰할 수 있으나 더 많은 공정 데이터 셋에 대한 패턴 확인 작업을 해야 하기 때문에 공정 데이터 셋의 인스턴스 수와 적절한 k값의 선정이 필요하다. 공정 데이터 셋의 크기 및 k의 값은 응용 분야에 맞게 적절하게 결정한다.



여기서  
 $C_i$ 는 생산 DB에서 생성된 i번째 인스턴스,  
 생산 공정은  $\dots, C_i, C_{i+1}, C_{i+2}, \dots, C_{i+m}, C_{i+m+1}, \dots$  순으로 진행,  
 한 공정 데이터 셋의 인스턴스 수는 m,  
 k는 연속 공정에서 이전 데이터 셋에 새로이 포함되는 인스턴스 수 ( $1 < k < m$ )  
 를 의미함

<그림 7> 연속 공정에서의 공정 데이터 셋의 구성

### 5.2 유사 정규패턴 추출

이 단계에서는 구성된 공정 데이터 셋에 대해

가장 유사한 정규패턴을 추출하게 된다. 가장 유사한 정규패턴은 공정 데이터 셋으로 부터 생성한 패턴과 저장된 정규패턴과의  $x^2$  통계량을 계산하여 가장 적은  $x^2$  통계량 값을 갖는 사례베이스내의 정규패턴을 유사 정규패턴으로 추출한다. 유사 정규패턴을 추출하기 위한 식은 다음과 같다.

$$x^2 = \min_q \left[ \max_j \left\{ \left( \frac{C_q - N_q}{N_q} \right)^2 + \left( \frac{(M - C_q) - (M - N_q)}{(M - N_q)} \right)^2 \right\} \right], \quad q = 1, \dots, n \tag{10}$$

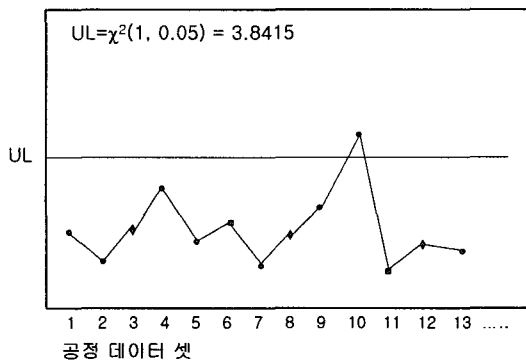
식(10)에서 사례 베이스에 n개의 정규패턴이 있을 때,  $C_q$ 는 q번째 정규패턴의 SOM 신경망으로 공정 데이터 셋의 출력패턴을 생성한 경우에 출력노드 j의 인스턴스 수를 나타내며  $N_q$ 는 q번째 정규패턴의 j 출력노드에 속한 인스턴스 수를 의미한다. M은 한 데이터셋의 인스턴스의 총 수를 나타낸다. 즉 공정 데이터 패턴과 q번째 정규패턴의  $x^2$  통계량을 각 출력노드에 대해 계산하여 가장 큰 값을 가진 출력노드의 수치를 q번째 정규분포를 기준패턴으로 했을 때의  $x^2$  통계량으로 고려한다. 모든 정규패턴에 대하여  $x^2$  통계량을 계산하여 가장 작은 수치를 나타내는 정규패턴을 현 공정 데이터 셋의 유사 정규패턴으로 추출한다.

### 5.3 다변량 공정관리도 작성

유사 정규패턴에 대한 공정 데이터 셋의  $x^2$  통계량 값이 정해진 유의수준에서의  $x^2$  값보다 크지 않은 경우 현 공정패턴은 유사 정규패턴과 같은 공정 상태라고 판정할 수 있다. 이를 관리도 형태로 그려보면 <그림 8>과 같은 상한선을 가

진 관리도로 나타낼 수 있을 것이다. <그림 8>은 공정 데이터 셋 13개에 대한 공정 상태를 나타내고 있으며 10번째 공정패턴의 경우 가장 유사한 정규패턴의  $\chi^2$  통계량 값이 상한선을 넘음을 알 수 있고 따라서 10번째 공정패턴은 현재 저장된 사례베이스내의 어떤 정규패턴과도 일치하지 않으므로 이상패턴일 가능성이 높다고 판단할 수 있다.

물론 관리상한선의 결정에 사용되는 유의수준은 응용 분야의 실험을 거쳐 가장 오류가 적은 유의수준을 선정하도록 한다. 그리고 정규패턴의 모양을 다르게 관리도에 표현함으로써 공정 상태의 패턴 변화를 표현할 수도 있을 것이다. 공정 상태가 정규패턴에 속하면 계속하여 생산 공정의 이상 유무 검출을 위한 공정 데이터 셋 구성 단계를 반복한다. 만약 사례베이스내의 정규패턴이 아닌 경우 이상 유무를 확인하는 유사 이상패턴 추출 단계로 진행한다.



<그림 8> 다변량 관리도의 예

#### 5.4 유사 이상패턴 추출

이 단계에서의 작업은 유사 정규패턴 추출 단계와 동일하게 진행한다. 다만 기준패턴으로 사

례베이스내의 이상패턴이 사용된다는 점만 다르다. 가장 유사한 이상패턴을 찾고 이상패턴에 대해 적합도 검정을 하여 이상패턴에 속한 경우 현 공정 상태는 밝혀진 이상 상태로 판정하고 공정 상태 시정 단계로 진행한다. 만약에 가장 유사한 이상패턴의 적합도 검정결과가 기각역에 속하면 현재의 공정패턴은 사례베이스내의 어느 패턴에도 속하지 않는다는 것을 의미한다. 이 경우 현 공정패턴에 대한 판정을 위해 판정확인 단계로 진행한다.

#### 5.5 판정확인

판정 확인단계에서는 공정 데이터 셋에 대해 이상 혹은 정상 상태 여부를 판정하기 위한 분석을 한다. 분석은 공정 데이터 셋에 포함되어 있는 각 변수에 대해 실시한다. 각 변수별로 어떤 이상이 존재하는지에 대해 기존의 단일 관리도를 이용하거나 신경망을 이용한 이상패턴 검출 방법(Hwang, 1993; Guh et al., 1999) 등이 적용될 수 있다. 어떤 경우든 생산공정의 엔지니어나 현장 전문가의 도움을 얻게 된다. 공정 데이터 셋에 포함되어 있는 변수가 많은 경우에는 변수의 우선 순위를 고려하여 이상 유무를 판단하는 것이 효율적이다. 변수의 우선 순위 결정을 위해서는 주성분 분석 같은 통계적 기법이나 혹은 신경망의 민감도 분석을 이용한 기법(강부식 & 박상찬, 2001) 등의 적용이 가능하다. 각 변수의 이상 유무의 결과 공정 데이터 셋은 이상패턴 혹은 정상 패턴으로 판정된다. 판명된 패턴은 현재 사례베이스에 존재하지 않는 패턴임으로 다음 단계에서 정상 패턴 혹은 이상패턴의 한 패턴으로 사례베이스에 추가한다. 또한 이상패턴으로 판명된 경우, 공정 이상 상태를 시정하기 위한 단계로

진행하여 필요한 조치를 취하게 된다.

### 5.6 새로운 정규 혹은 이상패턴 저장

현재 공정 데이터 셋을 SOM으로 학습하고 학습된 연결가중치와 SOM 출력패턴을 새로운 이상패턴의 유형으로 사례베이스에 추가한다.

## 6. 결론

현대의 제조 공정에서는 많은 품질특성치를 동시에 관리하는 것이 필요하나 통계적 다변량 관리도 기법을 적용하기에는 실무적으로 어려움이 많았다. 이 연구에서는 SOM과 적합성 검정을 이용하여 다변량 공정관리가 가능함을 실험적으로 제안하였다. 제안된 방법을 사용하면 많은 품질특성치를 공정 데이터 셋으로 구성하여 학습된 SOM을 이용하여 공정패턴을 생성하고, 생성된 패턴을 기준패턴과 적합성 검정하여 공정의 이상 유무를 판단할 수 있다. 따라서 적절한 기준패턴을 사례베이스에 저장하고 있는 경우 효과적으로 다변량 공정관리를 할 수 있다. 제안된 절차에 따라 초기 다변량 공정관리를 실시하는 환경에서는 많은 기준패턴을 준비할 수 없는 경우가 많으나 기준패턴이 점차 증가함에 따라 효과적인 관리가 가능하게 될 것이다. 실험을 통해 1종 오류를 점차 줄이면서 2종 오류는 아주 작게 유지함으로써 SOM과 사례기반추론이 결합된 절차가 다변량 공정관리를 위한 하나의 효과적인 방법론이 될 수 있음을 보였다.

그러나 제안된 방법에 따라 실제로 다변량 공정관리를 하기 위해서는 응용 분야 데이터의 특성에 따라 많은 실험이 추가적으로 이루어져야

한다. 실험을 바탕으로 패턴에 대한 정의가 선행되어야 하고 패턴을 생성하는 각 파라미터의 결정도 이루어져야 한다. 또한 이 연구는 주로 사례기반추론의 적용 가능성에 초점을 맞췄으나 추가적으로 좀 더 효율적인 패턴 저장 방법과 패턴 검출 방법 등에 대한 연구가 필요하다.

## 참고문헌

- 강부식과 박상찬, "신경망의 민감도 분석을 이용한 귀납적 학습기법의 변수 부분집합 선정", *한국지능정보시스템학회논문지*, 제7권 제2호 (2001), 51~63.
- 유성진, *원류관리에 종합품질경영 이론의 적용: SOM의 응용*, 한국과학기술원 석사학위논문, 1997.
- 홍창익, *인공신경망을 이용한 다변량 관리도의 이상 원인 해석*, 한국과학기술원 석사학위논문, 1995.
- Aamodt, A. and E. Plaza, "Case-Based Reasoning: Foundational Issues Methodological Variations and System Approaches", *AI Communications*, Vol. 7(1994), 39-59.
- Al-Ghanim, A., "An Unsupervised Learning Neural Algorithm for Identifying Process Behavior on Control Charts and a Comparison with Supervised Learning Approaches", *Computers and Industrial Engineering*, Vol. 32, No. 3(1997), 627-639.
- Bigus, J.P. and J. Bigus, *Constructing Intelligent Agents using JAVA*, 2<sup>nd</sup> Ed., John Wiley & Sons, NY, 2001.
- Guh, R.S., Y.C. Hsieh, "A Neural Network Based Model for Abnormal Pattern Recognition of Control Charts", *Computers & Industrial Engineering*, Vol. 36(1999), 97-108.
- Hotelling, H., "Multivariate Quality Control", in

- Eisenhart, Hastay and Wallis editors, *Techniques of statistical Analysis*, McGraw Hill, New York, NY, 1947.
- Hwang, H.B. and N. Hubele, "Back-Propagation Pattern Recognizers for X-bar Control Charts: Methodology and Performance", *Computers and Industrial Engineering*, Vol. 24, No. 2(1993), 219-235.
- Kang, B.S. and S.C. Park, "Integrated Machine Learning Approaches for Complementing Statistical Process Control Procedures", *Decision Support Systems*, Vol. 29(2000), 59-72.
- Kohonen, T., "Self-Organizing Map", *Proceedings of the IEEE* 78, No. 9(1990), 1464-1480.
- Kourti, T. and J.F. Macgregor, "Multivariate SPC Methods for Process and Product Monitoring", *Journal of Quality Technology*, Vol. 28, No. 4(1996), 409-428.
- Lowry, C.A., W.H. Woodall, C.W. Champ and S.E. Rigdon, "A Multivariate Exponentially Weighted Moving Average Control Chart", *Technometrics*, Vol. 34(1992), 46-53.
- Mason, R.L., C.W. Champ, N.D. Tracy, S.J. Wierda and J.C. Young, "Assessment of Multivariate Process Control Techniques", *Journal of Quality Technology*, Vol. 28, No. 4 (1997), 140-143.
- Pham, D.T. and E. Oztemel, "Control Chart Pattern Recognition using Learning Vector Quantization Networks", *International Journal of Production Research*, Vol. 32, No. 3(1994), 721-729.
- Watson, I., "Case-Based Reasoning is a Methodology not a Technology", *Knowledge-Based Systems*, 12 (1999), 303-308.
- Wierda, S.J., "Multivariate Statistical Process Control - Recent Results and Directions for Future Research", *Statistical Neerlandica*, Vol. 48 (1994), 147-168.





Abstract

## **Integrated Procedure of Self-Organizing Map Neural Network and Case-Based Reasoning for Multivariate Process Control**

Boo Sik Kang\*

Many process variables in modern manufacturing processes have influence on quality of products with complicated relationships. Therefore, it is necessary to control multiple quality variables in order to monitor abnormal signals in the processes.

This study proposes an integrated procedure of self-organizing map (SOM) neural network and case-based reasoning (CBR) for multivariate process control. SOM generates patterns of quality variables. The patterns are compared with the reference patterns in order to decide whether their states are normal or abnormal using the goodness-of-fitness test. For validation, it generates artificial datasets consisting of six patterns, normal and abnormal patterns. Experimental results show that the abnormal patterns can be detected effectively. This study also shows that the CBR procedure enables to keep Type 2 error at very low level and reduce Type 1 error gradually, and then the proposed method can be a solution for multivariate process control.

**Key Words:** Self-Organizing Map Neural Network, Case-Based Reasoning, Multivariate Process Control

---

\* Department of Management Information Systems, Mokwon University