
저자 (Authors)	윤재준, 박정술, 백준걸
출처 (Source)	대한산업공학회 추계학술대회 논문집 , 2010.11, 191-198(8 pages)
발행처 (Publisher)	대한산업공학회 Korean Institute Of Industrial Engineers
URL	http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE01960474
APA Style	윤재준, 박정술, 백준걸 (2010). SOM을 이용한 주기신호의 이상탐지 및 시각화. 대한산업공학회 추계학술대회 논문집, 191-198
이용정보 (Accessed)	한신대학교 211.187.***.147 2020/01/29 18:22 (KST)

저작권 안내

DBpia에서 제공되는 모든 저작물의 저작권은 원저작자에게 있으며, 누리미디어는 각 저작물의 내용을 보증하거나 책임을 지지 않습니다. 그리고 DBpia에서 제공되는 저작물은 DBpia와 구독계약을 체결한 기관소속 이용자 혹은 해당 저작물의 개별 구매자가 비영리적으로만 이용할 수 있습니다. 그러므로 이에 위반하여 DBpia에서 제공되는 저작물을 복제, 전송 등의 방법으로 무단 이용하는 경우 관련 법령에 따라 민, 형사상의 책임을 질 수 있습니다.

Copyright Information

Copyright of all literary works provided by DBpia belongs to the copyright holder(s) and Nurimedia does not guarantee contents of the literary work or assume responsibility for the same. In addition, the literary works provided by DBpia may only be used by the users affiliated to the institutions which executed a subscription agreement with DBpia or the individual purchasers of the literary work(s) for non-commercial purposes. Therefore, any person who illegally uses the literary works provided by DBpia by means of reproduction or transmission shall assume civil and criminal responsibility according to applicable laws and regulations.

SOM을 이용한 주기신호의 이상탐지 및 시각화

Fault Detection and Visualization of Cyclic Signals Using Self Organizing Map

윤재준, 박정술, 백준걸

고려대학교 산업경영공학과

junyee17@korea.ac.kr, dumm97@korea.ac.kr, jungeol@korea.ac.kr

Abstract

주기신호는 반도체 공정과 같은 고밀도, 고집적화된 제품 생산 공정 장비에서 발생하고 측정되는 대표적 데이터이다. 공정의 센서를 통해 측정되는 주기신호의 이상탐지를 위해서 주기신호를 구간으로 나누어 구간별로 SPC 차트적용 시키는 방법, 각 시점마다 측정되는 값을 하나의 변수로 고려하여 Hotelling's T^2 , PCA, PLS 등과 같은 다변량 통계 분석을 적용 시키는 방법들이 제시되어 왔다. 이러한 방법들은 다양한 특성을 가지는 주기신호를 분석하고 이상탐지 하는데 많은 한계점을 가진다.

이에 본 논문은 SOM(Self Organizing Map)을 이용하여 주기신호의 특성을 반영한 시각화와 공정의 이상을 탐지하는 방법을 제안한다. 제안하는 알고리즘은 기존 SOM에서 맵 상 하나의 winning node로 하나의 데이터를 표현하는 방식과 달리 맵 상 모든 노드를 이용하여 데이터를 표현한다. 이를 바탕으로 주기신호를 시각화시키고 새로 표현된 데이터의 특징(feature)을 기반으로 Logistic regression을 적용시켜 이상을 탐지 한다. 다양한 이상 상황을 가진 반도체 공정 주기신호 데이터를 사용하여 제안한 이상탐지 성능을 평가하였다.

Key words : Cyclic Signal, Visualization, SOM, Fault Detection

1. 서론

반도체와 같이 고밀도, 고집적화 된 제품 생산을 위하여 세밀하고 정확한 작업을 가능케 하는 공정 장비가 도입되고 있다. 진보된 공정 장비는 다양한 데이터를 발생시킨다. 공정 장비의 발달과 더불어 공정 센서의 발달로 인해 공정 장비가 발생시키는 데이터 측정 또한 용이해졌다. 그 결과 각 공정에서 다양한 형태의 데이터가 측정되고 있다. 특히 주기신호는 반도체 공정의 Chemical Vapor Deposition(CVD) 공정에서와 같이 실제 많은 공정에서 발생하고 측정되는 대표적 데이터이다.

주기신호는 <Figure 1>과 같이 각 시점에 측정되는 값들이 하나의 주기를 기준으로 특정 패턴을 가지고 반복적으로 발생하는 데이터를 뜻한다. 공정 상황이 정상일 경우 정상 패턴을 유지하지만 이상이 발생하였을 때 정상과 다른 패턴을 보이며 또한 이상 원인에 따라 각기 다른 특징을 가진 이상 패턴을 보인다. 따라서 주기신호의 특징을 파악하고 이상을 탐지하는



Figure 1. 주기신호의 예

것은 공정을 효과적으로 제어하고 관리하기 위해 매우 중요한 과제이다.

주기신호의 이상 탐지를 위해 다양한 방법들이 제안 되어 왔다. 현 공정에서 널리 사용되는 방법은 주기신호의 특징에 따라 구간을 나누고 구간별로 데이터의 분산을 추정하고 유의 수준(α)내의 관리 상, 하한선을 적용하여 Shewhart 차트, EWMA 차트 등의 Statistical Process Control (SPC) 차트로 관리하는 방식이다. 이러한 방식은 변수의 개수(n)가 증가할수록 $(1 - \alpha)^n$ 의 비율로 정상신호를 이상으로 판단할 확률이 증가 한다. 또한 <Figure 2>와 같이 이상 발생 확률이 큰 구간 중 하나인 신호가 급격히 변하는 곳에서의 분산이 크기 때문에 넓은 관리 한계선을 가지게 되어 이상탐지 성능이 떨어지는 한계점을 가진다.

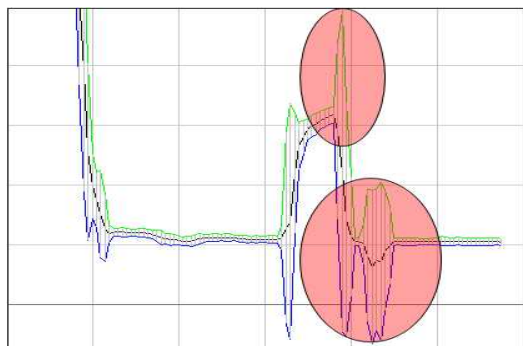


Figure 2. 3 σ -SPC Chart

다른 방법으로는 Hotelling's T^2 , PCA(Principal Component Analysis), PLS(Partial Least Square) 등과 같은 다변량 통계 분석을 적용 시키는 방법들이 제시되어 왔다. PCA, PLS는 변수를 변환하는 과정을 거치기 때문에 공정의 이상을 탐지하더라도 그 원인을 찾는 것이 힘들고 복잡하다. 따라서 이상 타입의 구분이 어렵다.

위에서 언급한 여러 이상탐지 방법의 단점들을 해결하고자 본 연구에서는 SOM을 기반으로 하는 주기신호 이상탐지 알고리즘과 이상 타입별로 구분지어 관리할 수 있는 주기신호 시각화 방법을 제안한다.

본 논문의 구성은 다음과 같다. 2장 본론에서는 2.1장 제안한 알고리즘에서 사용될 SOM의

설명과 적용 방식, 2.2장 제안된 SOM를 적용하여 주기신호의 시각화 방법, 2.3장 제안된 SOM을 적용한 이상탐지 알고리즘을 제시한다. 3장에서는 실험 및 결과 분석을 통해 제안하는 알고리즘의 효용성을 보이고 4장에서는 결론을 기술한다.

2. 본론

2.1 SOM(Self Organizing Map)

SOM은 인공신경망(Artificial Neural Network)의 competition learning 방식 중 하나의 방식이다. 다차원의 데이터를 낮은 차원(2차원 또는 3차원)의 맵 상으로 표현하는 다차원 데이터의 시각화에 매우 유용한 알고리즘이다.(T. Kohonen, 1990) 본 논문에서는 주기신호의 시각화를 위해 데이터를 2차원의 맵으로 표현한다. 다음은 SOM을 이용하여 주기신호를 학습하는 과정이다.

2차원 맵상 M개의 노드(node)는 각각 $w_i = (w_{i1}, w_{i2}, w_{i3}, \dots, w_{in})$, $i = \{1, 2, 3, \dots, M\}$ 의 weight 벡터를 가지고 있다. 길이가 n인 K개의 입력 벡터 $x_k = (x_{k1}, x_{k2}, x_{k3}, \dots, x_{kn})$, $k = \{1, 2, 3, \dots, K\}$ 가 2차원 맵 상 하나의 노드에 mapping 되는데 그 노드를 결정하는 방식은 식(1)과 같다. 입력 벡터는 길이가 일정하게 보정된 주기신호를 뜻한다.

$$c_k = \arg \min \| X_k - W_i \| \quad (1)$$

c_k 는 입력 벡터 x_k 가 mapping되는 2차원 맵상의 노드 c, w_i 는 노드의 weight 벡터이다. 즉, M개의 노드 중 입력 벡터 x_k 와 euclidian distance가 가장 가까운 노드(winning node)에 입력 벡터가 mapping된다. t시점의 하나의 입력 벡터가 winning node로의 입력이 끝난 후 식 (2)와 같은 방식으로 i번째 노드의 t+1시점의 weight 벡터를 update 한다.

$$W_i(t+1) = W_i(t) + \alpha(t) \cdot h_{ci}(t) \cdot [X(t) - W_i(t)] \quad (2)$$

$\alpha(t)$ 는 시간에 따라 감소하는 learning rate에

관한 함수이다. $h_{ci}(t)$ 는 노드 c 와 이웃하는 노드 i 의 update 비율을 결정하는 함수이다. 즉, winning node의 weight 벡터만을 update 하는 것이 아니라 이웃하는 노드까지 업데이트 하는 것을 의미한다. $h_{ci}(t)$ 함수에는 bubble, gaussian, ep, cutgauss 와 같은 함수들이 있다. 모든 데이터의 winning node가 결정되면 학습은 종료된다. SOM에 의해 학습된 데이터는 특징 별로 분리된 위치에서 클러스터링(clustering)되어 2차원 맵 상에 표현된다.

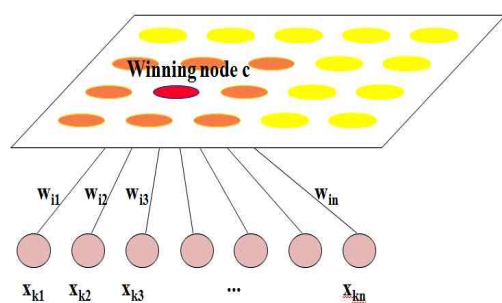


Figure 3. 4×5 SOM의 예

학습과정이 끝난 후 데이터를 SOM 상에 적용하는 과정에서 기존 SOM은 하나의 데이터를 2차원 평면 상 하나의 winning node로 표현한다. 본 논문의 목적은 주기신호를 시각화시키는 것 뿐 만 아니라 이상탐지 알고리즘까지 구현하는 것이다. 다양한 정보를 가진 주기신호를 2차원 맵 상 하나의 점으로만 표현 하는 것은 주기신호의 특성을 충분히 반영하지 못하여 이상탐지 성능을 저하시킨다.

따라서 본 논문에서는 SOM에 mapping되는 하나의 데이터를 winning node만으로 표현 하지 않고 전체 맵 상 모든 노드의 weight 벡터와 데이터 사이의 거리로 데이터를 설명하는 방식을 제안한다. 즉, 새로운 데이터 x_i 가 주어졌을 때 식(3)에서 표현되는 M 개의 d_{ti} 가 데이터의 새로운 특질이 된다.

$$d_{ti} = \|X_t - W_i\| \quad (3)$$

$d_i = (d_{t1}, d_{t2}, \dots, d_{ti})$ 는 입력 벡터 x_i 와 i 노드의

weight 벡터 w_i 와의 거리를 뜻한다.

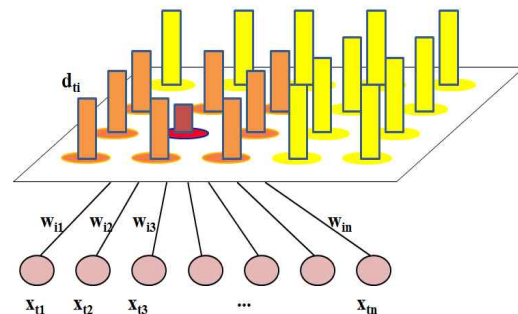


Figure 4. 제안하는 데이터 표현방식

2.2 SOM을 기반으로 하는 주기신호의 시각화

주기신호는 하나의 주기를 가지고 특정 패턴을 반복적으로 발생시키는 공정 데이터이다. 이러한 주기 신호는 이상 원인에 따라서 각기 다른 패턴을 띄게 된다. 특히 신호의 변동 폭이 큰 구간에서 실제 패턴과 다르게 변하는 타입의 이상이 많다. 예를 들어 공정의 압력을 나타내는 주기신호를 보면 압력이 일정히 유지되는 구간이 있고, 반면에 압력이 갑자기 증가하는 구간이 있다. 압력이 일정하게 유지되는 구간은 공정의 제어가 비교적 쉽기 때문에 이상이 많이 발생하지 않는다. 하지만 갑자기 증가하는 구간에서는 공정 설비가 정확하게 공정을 제어를 하지 못하게 되는 경우가 있고 그로 인해 공정의 압력이 갑자기 튀어 오르게 되는 이상패턴이 발생한다.

기존 SPC 차트를 통해 주기신호를 관리하게 되면 서론에서 설명한 것처럼 이러한 이상 패턴에 대한 감지 능력이 떨어지고 여러 이상 타입을 구분지어 관리하지 못한다.

따라서 본 논문은 2.1장에서 설명한 SOM을 기반으로 하는 주기신호 데이터의 시각화 방법을 다음과 같이 제안 한다. 정상 주기신호와 여러 타입의 이상 주기신호를 SOM을 통해 학습시킨다. 주기신호를 학습시킨 맵은 각 노드마다 weight 벡터를 가진다. SOM의 학습 특성상 weight 벡터와 입력된 주기신호의 거리가 가까운 노드에 주기신호가 mapping되고 weight가 입력된 주기신호와 비슷한 모습을 띄게 update 되는 과정을 거치기 때문에 학습과정이 진행될수록

SOM의 각 노드가 가지는 **weight** 벡터는 입력되는 정상, 이상 주기신호들의 특성을 반영하는 모습으로 점차 변하게 된다. 학습과정이 끝나면 SOM은 정상, 여러 타입의 이상 주기신호 **cluster**로 구분 되어진 맵이 형성된다.

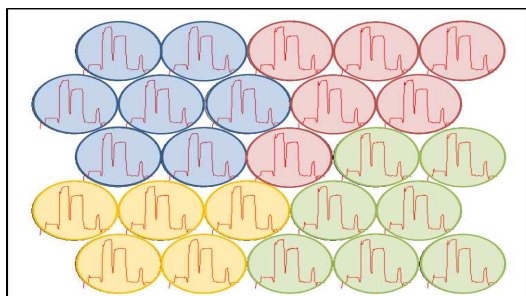


Figure 5. 주기신호를 학습한 SOM 예

주기신호를 2.1장에서 설명한 방식으로 표현하기 위해 학습이 끝난 주기신호 x_k 를 다시 맵에 **mapping** 시키면 2.2장에서 설명한 것과 같이 **winning node**의 d_k 값이 제일 작은 것을 기준으로 이웃하는 노드의 d_k 값이 점차 커지는 <Figure 6>과 같은 3D plot으로 주기신호 x_k 가 표현된다.

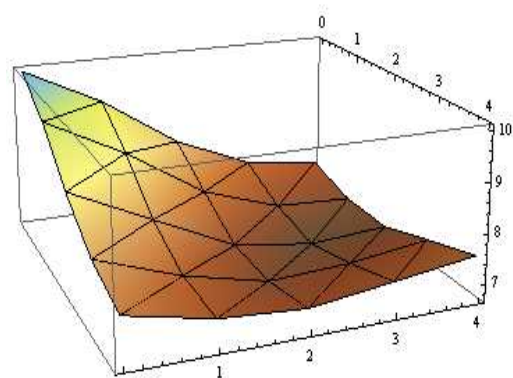


Figure 6. 주기신호의 3D plot의 예

학습과정에서 사용된 주기신호 x_k 는 정상 주기신호, 타입별 이상 주기신호에 따라 서로 다른 3D plot을 가진다. 새로운 주기신호 x_i 가 들어왔을 때 d_i 값에 따라 새로운 3D plot이 생성된다. 이 3D plot은 학습된 3D plot과 비교 과정을 통해 정상 또는 어떤 타입의 이상인지 판별이 가능하다. 또한 여러 타입의 이상의 패턴을 동시에 가지고 있는 경우에도 3D plot으로

확인 가능한 장점을 보인다. 주기신호 x_i 의 특질 d_i 를 기반으로 하는 이상탐지 알고리즘에 대해 2.3장에서 기술한다.

2.3 SOM을 기반으로 하는 이상탐지 알고리즘

2.1장에서 설명한 SOM의 학습이 끝나고 2차원 맵 상 모든 노드의 w_i 가 결정된다. 다시 학습주기신호 x_k 를 SOM상에 **mapping** 시키면 x_k 의 새로운 M개의 d_k 가 계산된다. d_k 가 의미하는 것은 2차원 맵 상의 하나의 점을 3차원 모습으로 바꾸어 주는 것이다. 따라서 정상 주기신호와 여러 타입의 이상 주기신호가 3차원의 모습으로 표현되어 진다. 학습된 정상 주기신호의 **winning node**와 이상 신호의 **winning node**는 각각 구분되어 클러스터링 된다. 학습된 주기신호의 d_k 를 계산해보면 SOM의 학습 규칙에 따라 **winning node**의 d_k 가 가장 작은 것을 중심으로 이웃하는 노드의 d_k 가 점차 증가하는 것을 알 수 있다. 따라서 주기신호의 새로운 특질 d_k 는 주기신호의 타입에 따라 각각 다른 **winning node**를 중심으로 가장 낮고 이웃 노드의 d_k 가 증가하는 특징적인 모습을 갖게 된다. 따라서 새로운 주기신호의 d_k 값과 학습된 주기신호의 d_k 값들과의 차이를 통해 정상 또는 어떤 타입의 이상 주기신호인지 판단할 수 있다. 주기신호의 새로운 특질 d_k 의 차이를 판단하는 기준이 되는 정확한 척도를 위해 다음과 같이 **logistic regression** 적용한다. (Hosmer, 2000)

x_k 의 M개의 새로운 feature d_k 와 클래스 정보를 포함하여 새로운 데이터 셋을 만든다. 길이가 M인 d_k 를 변수로 하는 **logistic regression**을 수행하여 모델을 만든다.

$$f(z) = \frac{e^z}{e^z + 1} \quad (4)$$

$$z = \beta_0 + \beta_1 D_{k1} + \beta_2 D_{k2} + \dots + \beta_M D_{kM} \quad (5)$$

식(4,5)는 **logistic regression**의 모델을 나타낸 것이고 t 는 주기신호의 타입, β 는 **Multiple Logistic regression**의 계수이다. 이 계수는 **MLE (Maximum Likelihood Estimation)**을 이용하여

추정한다. 정상 주기신호와 타입별 이상 주기신호의 특징을 사용하여 각각 로지스틱 모델 생성 후, 테스트 주기신호가 들어왔을 때 각 모델에서의 확률이 가장 큰 타입으로 테스트 주기신호의 타입을 결정한다.

logistic regression 은 정규 성 가정을 하지 않기 때문에 정규 분포하지 않는 주기신호 분석에 적합하다. 또한 테스트 주기신호의 타입을 확률 값으로 제공해주므로 두 개의 이상 패턴이 섞여 있는 주기신호가 들어오더라도 각 이상 주기신호 타입일 확률이 높고 비슷하게 할당되므로 두 개 패턴이 섞여 있다는 것을 확인할 수 있는 장점이 있다. 따라서 3차원 공간으로 표현되는 정상 주기신호 데이터, 이상 주기신호 데이터와 테스트 주기신호 간의 닮은 정도를 파악하여 정상, 이상 여부를 판단하는 기준을 제공한다.

3. 실험 및 결과 분석

3.1 반도체 공정 주기신호 데이터 실험

제안한 시각화와 이상탐지 알고리즘의 성능 확인을 위하여 CVD(Chemical Vapor Deposition) 공정의 OES(Optical Emission Spectroscopy) 센서를 통해 확보된 주기신호를 바탕으로 시뮬레이션을 통해 정상 주기신호와 3개의 타입의 이상 신호를 가공하였다. 각 타입별 주기신호를 특정 분산을 가정하고 1000개씩 생성하였고 실험 전 전처리 과정을 통해 주기신호의 길이를 256으로 보정하였다. 각 주기신호는 <Figure 7, 8, 9, 10>과 같은 특징을 가진다.

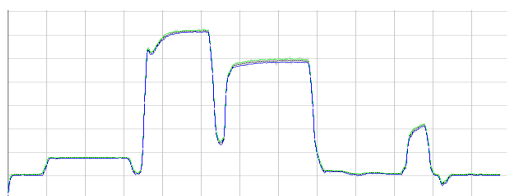


Figure 7. 정상 주기신호

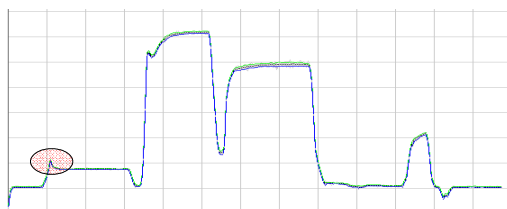


Figure 8. 이상 주기신호 Type 1

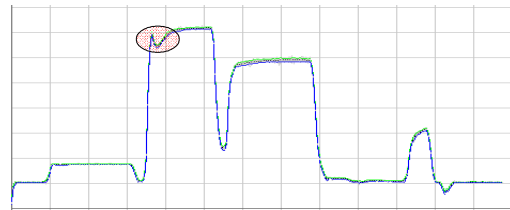


Figure 9. 이상 주기신호 Type 2

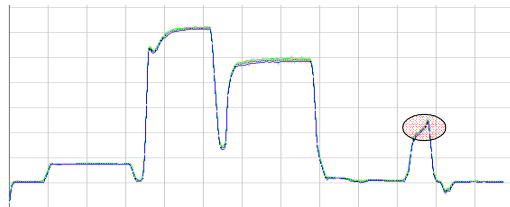


Figure 10. 이상 주기신호 Type 3

이상 주기신호 타입 1, 2, 3은 실제 공정에서 많이 발생하고 있는 이상 중에 신호가 급격히 변하는 구간에서 설비가 제어 하지 못하고 신호가 튀는 현상을 반영하여 생성하였다.

각 1000개씩 4종류의 주기신호 중 80%를 random 하게 선택하여 SOM 학습에 사용하고 나머지 20%를 주기신호 테스트 데이터로 사용하였다. 학습 데이터로 생성한 SOM의 weight 벡터들이 주기신호의 특징을 잘 반영하는지 확인하였고 학습 주기신호의 d_k 를 구하여 이를 바탕으로 한 새로운 형태의 주기 신호의 시각화가 각각 주기신호별로 어떻게 다른지 확인하는 실험을 하였다.

실험 과정에서 SOM의 효율성을 보여주기 위해 타입1, 3의 이상신호 패턴을 섞은 새로운 이상신호도 첨가하여 새로운 형태의 주기 신호의 시각화가 어떤 형태를 보이는 테스트 하였다. 또한 학습 데이터의 특징 d_k 와 클래스 정보로 logistic regression을 수행하여 각각 모델을 만들고 테스트 주기신호의 d_k 를 모든 모델에 적용하여 테스트 주기신호가 어떤 타입의 주기 신호인지에 대한 확률을 구함으로써 제안 알고리즘의 이상 탐지 성능을 확인해 보았다.

제안하는 알고리즘의 특성상 SOM 학습이 매우 중요하므로 맵의 크기, 반복횟수, $\alpha(t)$, $h_{ci}(t)$ 의 SOM 모수들은 반복 실험을 통해 결정하였다.

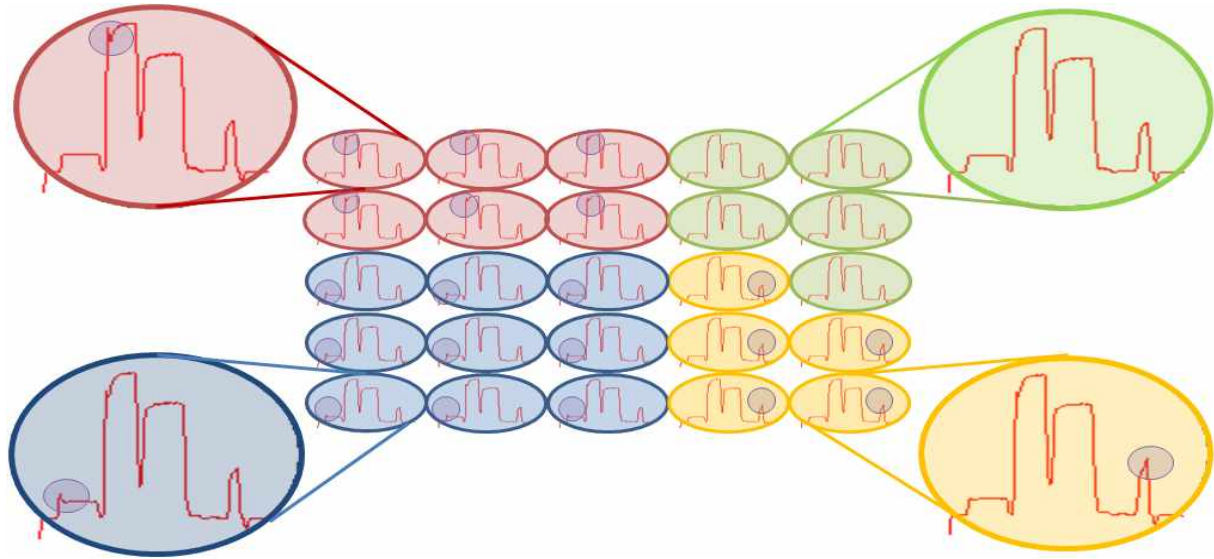


Figure 11. 학습 주기신호를 5×5 SOM에 학습 시킨 후 각 노드의 weight 벡터의 모습

3.2 결과 분석

학습 주기신호 데이터를 5X5 SOM으로 학습한 결과는 <Figure 11>과 같다. 학습된 SOM의 weight 벡터를 살펴보면 주기신호의 특징을 잘 반영한 모습을 보이고 있는 것을 알 수 있다. 녹색 노드가 정상 주기신호, 파란색 노드가 이상 주기 신호 타입 1, 빨간색 노드가 이상 주기 신호 타입 2, 노란색 노드가 이상 주기 신호 타입 3을 나타낸다. 녹색 노드와 노란색 노드의 경계 부분의 노드의 weight 벡터가 정상과 이상 주기 신호 타입 3 중간 형태의 모습을 띄고 있는 것을 볼 수 있는데 이는 시뮬레이션으로 생성한 주기신호의 분산에 의해 나타나는 다양한 형태의 주기신호 중 정상 주기 신호로 들어온 것 가운데 이상 주기 신호 타입 3의 모습과 유사한 주기신호들과 주기 신호 타입 3으로 들어온 것 중 정상 주기 신호 타입과 비슷한 모습을 띤 주기신호의 영향으로 경계부근에 중간형태의 주기신호 모습으로 weight 벡터가 학습된 것을 의미한다. 이렇게 경계 노드에 매핑 되는 새로운 데이터는 logistic regression을 통해 구해지는 확률로 어떤 주기신호인지 판별할 수 있다.

학습 주기신호의 d_k 를 구하여 이를 바탕으로 한 새로운 형태의 주기 신호의 시각화는 예는 <Figure 12,13,14>와 같다.

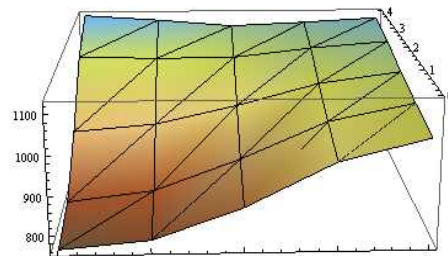


Figure 12. 이상 주기신호 타입 1

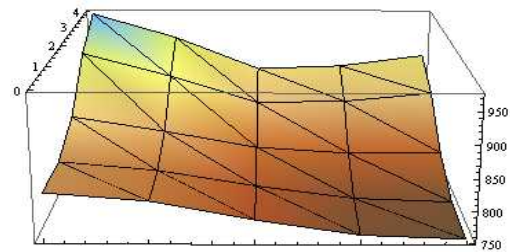


Figure 13. 이상 주기신호 타입 3

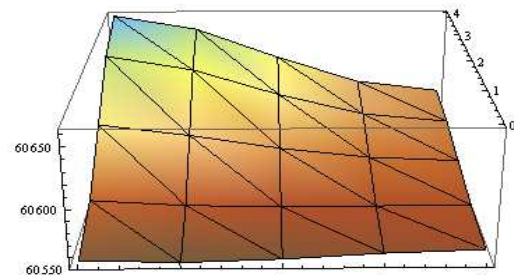


Figure 14. 이상 주기신호 1,3의 패턴을 가진 새로운 주기신호

<Figure 12,13>을 보면 각 주기신호 타입마다 d_k 값이 winning node에서 최소가 되고 주위로 갈수록 증가하는 모양을 나타낸다는 것을 알 수 있다. 따라서 새로운 주기신호가 들어 왔을 때 3d plot을 보고 시각적으로 어떤 타입의 주기신호인지 확인할 수 있다. 또한, <Figure 14>는 이상 주기신호 타입1, 3이 가지는 패턴을 동시에 가지고 있는 새로운 주기신호를 시각화 한 것이다. 특이한 점은 모든 노드와의 d_k 의 절대 값은 상당히 크지만 3d plot 모양으로 확인해 보면 이상 주기신호 타입1과 타입3의 winning node 부근에서 상대적으로 작은 값을 가지는 것으로 나타난다. 따라서 새로운 주기신호가 이상 주기신호 타입1, 3이 가지는 패턴을 둘 다 가지고 있다는 것을 시각적으로 확인할 수 있다.

제안하는 알고리즘을 사용하여 전체 주기신호의 80%를 학습 주기신호로 SOM에 학습시키고 logistic regression 모델을 만들어 나머지 20%의 테스트 주기신호에 적용하여 이상 탐지 성능을 측정한 결과는 다음 [Table.1]과 같다.

[Table.1] 이상 탐지 알고리즘 실험 결과

예측 실제	정상	타입1	타입2	타입3
정상	165	3	9	23
타입1	1	173	12	14
타입2	3	8	185	4
타입3	18	9	5	168

위 실험 결과는 각 타입에 속할 확률 중 가장 높은 곳에 테스트 주기신호를 할당하였다. 실험 결과 평균 86.4%의 이상 탐지 능력을 보인다. 하지만 확률적으로 어떤 타입의 주기신호인지 애매한 값들이 발생하는 상황이 있으므로 관리자의 역치 설정에 따라 변할 수 있는 여지가 있다.

이상 주기신호 타입1, 3의 패턴을 동시에 가지고 있는 새로운 테스트 주기신호에 logistic regression을 적용한 결과는 다음 [Table.2]와

같다.

[Table.2] 새로운 타입의 주기신호 실험 결과

	정상	타입1	타입2
타입1,3 혼합	28.8%	74.3%	68.2%

위의 실험결과는 정상 주기신호와 이상 주기신호 타입1의 logistic regression 모델과 정상 주기신호와 이상 주기신호 타입3의 logistic regression 모델을 타입1,3의 패턴이 혼합되어 있는 새로운 테스트 주기신호 200개에 적용하여 각 타입에 속할 확률을 구한 것이다. 두 개의 이상 주기신호 타입에서 모두 높은 확률을 보인다. 따라서 이상 주기신호 타입1, 3의 패턴을 동시에 가지고 있다는 것을 확률로 확인할 수 있다. 두 패턴을 동시에 가지는 이상 주기신호에 대하여 <Figure 11>은 시각적 정보를 제공해 주는 것이고 위의 실험은 logistic regression을 통해 확률 정보를 제공해 주는 것이다.

4. 결 론

본 논문은 주기신호의 이상 탐지와 시각화를 위하여 SOM을 기반으로 한 알고리즘을 제안하였다. 실험 결과에서 나타나듯이 주기신호의 시각화 부분에서 상당히 의미 있는 결과를 얻을 수 있었다. 관리자가 주기신호를 모니터링 할 때 어떤 타입의 이상 신호인지 판단 가능하도록 1차적으로 시각적 정보를 제공하고 2차적으로 확률 값을 제공해 주므로 빠른 판단을 내릴 수 있게 도움을 준다.

실험에 사용된 주기신호는 CVD공정의 OES 센서를 통해 측정된 주기신호를 바탕으로 시뮬레이션을 통해 생성하였다. 따라서 이상탐지 알고리즘의 결과는 시뮬레이션 과정에서 주기신호를 생성할 때 분산 설정에 상당한 영향을 받았다. 예를 들어 정상 주기신호의 분산을 크게 주고 주기신호를 생성하면 많은 이상 주기신호가 섞여있는 데이터가 정상 주기신호로 생성되어 이상탐지 성능에 큰 영향을 끼치게

된다. 따라서 실제 공정 데이터를 가지고 실험을 하면 좀 더 정확한 성능 평가가 가능할 것이라 생각된다.

또한 과거 주기신호 데이터를 바탕으로 정확한 주기신호의 이상 타입 정보를 학습해야 하는 단점이 있다. 의도치 못한 다른 타입의 주기신호가 발생했을 때 그 주기신호와 가장 가까운 형태를 지닌 주기신호로 구분되는 단점이 존재한다. 하지만 의도치 못한 주기신호의 타입은 결정 하지 못하지만 이상여부에 대한 정보는 제공해 줄 수 있다.

제한한 알고리즘은 SOM의 학습 결과에 큰 영향을 받는다. 따라서 SOM을 잘 학습하기 위해서 이상 데이터의 특징을 충분히 반영해야 하므로 이상 데이터의 양이 많이 필요하다. 또한 SOM의 학습 시 세팅하는 적정 모수들은 반복 실험을 통하여 구해야 하는 한계점이 있다.

추후 연구과제는 SOM의 적절하게 모수를 결정할 수 있는 방법을 찾는 연구와 여러 이상 주기신호 타입의 원인을 찾는 연구가 필요하다.

감사의 글

본 과제는 정보통신산업진흥원의 SW공학 요소기술 개발과 전문 인력 양성사업의 결과물임을 밝힙니다.

참고문헌

Hosmer, DW., Lemeshow, S. (2000), Applied Logistic Regression, John Wiley & Sons, New York, NY.

Kohonen, T. (1990), The Self-Organizing Map, Proceedings of the IEEE

Montgomery, Douglas C. (2001), Introduction to Statistical Quality Control, 5th Edition, John Wiley & Sons, New York, NY

Hostie T., Tibshirani R., Friedman J. The Elements of Statistical Learning : Data Mining,

Inference, and Prediction, Springer, New York

Laaksonen, Jorma T., Markus Koskela J., Erkki Oja (2004), Class distributions on SOM surfaces for feature extraction and object retrieval, Neural Networks 17(2004) 1121-1133

Liu Y., Weisberg Rober H., Mooers Christopher N. K. (2006), Performance evaluation of the self-organizing map for feature extraction, Journal of geophysical research, VOL. 111, C05018, doi:10.1029/2005JC003117

JaeHyun Lee, JiHyun Kim, JiBin Hwang, Sung-shick Kim (2007) A Study on Fault Detection of Cycle-based Signals using Wavelet Transform, 한국시물레이션학회 논문지 Vol 16, No. 4, pp 13-22