

웨이블릿을 이용한 주기 신호 데이터의 이상 탐지에 관한 연구

이재현¹ · 김지현^{2*} · 황지빈¹ · 김성식¹

A Study on Fault Detection of Cycle-based Signals using Wavelet Transform

JaeHyun Lee · JiHyun Kim · JiBin Hwang · Sung-Shick Kim

ABSTRACT

Fault detection of cycle-based signals is typically performed using statistical approaches. Univariate SPC using few representative statistics and multivariate analysis methods such as PCA and PLS are the most popular methods for analyzing cycle-based signals. However, such approaches are limited when dealing with information-rich cycle-based signals. In this paper, process fault detection method based on wavelet analysis is proposed. Using Haar wavelet, coefficients that well reflect the process condition are selected. Next, Hotelling's T^2 chart using selected coefficients is constructed for assessment of process condition. To enhance the overall efficiency of fault detection, the following two steps are suggested, i.e. denoising method based on wavelet transform and coefficient selection methods using variance difference. For performance evaluation, various types of abnormal process conditions are simulated and the proposed algorithm is compared with other methodologies.

Key words : Cycle-based Signal, Wavelet Transform, Haar Wavelet, Coefficient Selection

요약

주기 신호 데이터를 가지는 공정의 이상 탐지를 위해 대표값을 사용하는 단변량 SPC 차트나 PCA, PLS 등과 같은 다변량 통계방법들이 사용되고 있다. 이러한 방법들은 주기 신호 데이터의 다양한 정보를 분석하는데 한계가 있다. 본 연구에서는 Haar 웨이블릿 변환을 이용하여 주기 신호의 형태를 반영하는 웨이블릿 계수를 구하고, 이 계수들에 SPC 차트를 적용하여 공정 이상여부를 탐지하였다. 본 논문에서는 보다 효율적인 이상 신호 탐지를 위해 웨이블릿을 이용한 잡음 제거 기법과 Haar 웨이블릿 계수의 분산 차이를 이용한 중요 계수 선택 방법을 제안하였다. 다양한 이상 상황에 대하여 시뮬레이션을 통하여 제안한 알고리즘의 효율성을 확인하였다.

주요어 : 주기 신호 데이터, 웨이블릿 변환, Haar 웨이블릿, 중요 계수 선택

1. 서론

1.1 연구의 배경 및 필요성

최근 데이터 처리 기술이 급격하게 발전함에 따라, 공정에서는 이전에 측정 불가능했던 데이터들의 측정이 가능해졌다. 이로 인하여 반도체 공정과 같이 정밀한 제품을 가공하는 공정에서는 보다 많은 공정 데이터를 얻게 되었고, 얻어진 데이터를 통하여 세밀한 공정 분석이 가능해 지고 있다. 특히 주기성을 가지고 특정한 패턴을 가진 데이터(이하 주기 신호 데이터)는 기술 발전으로 인한

* 본 과제는 교육인적자원부, 산업자원부, 노동부의 출연금으로 수행한 최우수실험실지원사업의 연구결과입니다.

2007년 9월 9일 접수, 2007년 10월 19일 채택

¹⁾ 고려대학교 산업시스템정보공학과

²⁾ 고려대학교 반도체기술연구소

주 저 자 : 이재현

교신저자 : 김지현

E-mail: jihyunk@korea.ac.kr

여 이전보다 데이터 획득이 용이해 졌다. 자동차 공정에서의 스탬핑(stamping) 공정, 스폿 용접(spot welding) 공정, 엔진 조립 공정 등과 반도체 공정에서의 포토 공정, 화학기상증착 공정 등과 같이 실제 많은 제조 공정에서 주기 신호 데이터가 발생하고 있다. 주기 신호 데이터는 한 주기 동안에 각 시점별로 측정되는 데이터로, 공정 진행 상태에 대한 다양한 정보를 포함하고 있다.

일례로 그림 1의 welding force 주기 신호 데이터를 살펴보면^[1], 이 주기 신호 데이터는 한 주기가 세 개의 구간으로 구분되어 각 구간별로 다른 작업이 진행된다. 매 구간마다 다른 형태의 패턴이 형성되고, 이상(fault)이 발생하는 원인과 형태 또한 구간별로 다르게 발생한다. 이와 같은 공정에서는 한 주기 안에 많은 정보를 포함하고 있는 주기 신호 데이터를 효과적인 분석을 통해 공정의 이상을 탐지하고 이상의 원인을 파악하여 제품의 품질을 높일 수 있는 방법이 필요하다.

1.2 기존 연구의 고찰

1.2.1 전통적인 주기 신호 이상탐지

주기 신호 데이터에 대한 전통적인 분석 방법은 신호를 표현하는 대표 통계치를 단변량 SPC(Univariate Statistical Process Control) 차트에 적용하여 분석하는 방법이다. 이 방법은 주기 신호 데이터의 전체 평균이나 최대값, 최소값 등을 대표값으로 정하고, 이 값을 사용하여 공정의 이상을 탐지하였다. 이 방법은 탐지하기 어려운 이상요인이 많고 이상이 탐지되어도 원인 분석에 어려움이 있다.

이를 극복하는 방법으로 데이터가 관찰되는 각 시점을 하나의 변수로 보고 분석하는 방법이 제안되었다. 각 시점 별로 Shewhart 차트나 EWMA, CUSUM 차트를 사용하여 공정의 이상을 탐지하였다. 이 방법은 한 주기의 데이터가 길 경우 관리해야 할 차트가 많아져 비효율적이다. 예를 들어 20개의 공정에 측정되는 길이가 250개이고 공정당 4개의 변수를 측정한다면 20000개의 차트가 필요하게 된다. 또한 변수들 간의 상관관계가 고려되지 않기 때문에 탐지 못하는 불량률이 대량으로 발생한다.

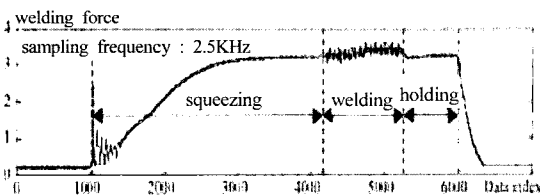


그림 1. Welding force 주기 신호 데이터

이와 같은 문제를 해결하기 위해 다변량 통계 분석 기법이 도입되었다. Hotelling's T^2 와 PCA(Principal Component Analysis), PLS(Partial Least Squares)등이 대표적인 방법이다^[2,3]. PCA나 PLS는 잠재 변수(Latent Variable)를 산출하여 차원을 축소하고 잠재 변수를 이용하여 분석하는 방법이다. 잠재 변수들은 연관성을 갖는 변수들의 선형조합이기 때문에 잠재 변수에 이상이 발생했을 경우, 실제 원인이 되는 변수를 찾는 과정이 평이하지 않다. 또한 이 방법들은 모인 샘플의 수보다 데이터의 길이가 클 경우 공분산을 계산할 수 없기 때문에 방법 적용이 불가능할 수 있다.

1.2.2 웨이블릿을 이용한 주기 신호 이상탐지

최근에는 주기 신호 데이터의 분석을 위해 신호 처리 분야에서 각광을 받고 있는 웨이블릿 변환(Wavelet Transform)을 활용한 연구가 진행되고 있다. 웨이블릿 변환은 다중 해상도 분석(Multiresolution Analysis)이 가능한 특성 때문에 많은 정보를 가진 주기 신호 데이터의 분석에 적합하다. 이 방법은 원 신호를 웨이블릿 변환을 적용하여 웨이블릿 계수들로 feature를 형성하고 이 feature들을 이용하여 이상을 탐지한다.

Bakshi(1998)는 웨이블릿 계수를 구한 후 MSPCA (Multi-scale PCA)를 사용하여 공정의 이상을 탐지하였다^[4]. Koh 등(1999)은 웨이블릿 변환을 통하여 웨이블릿 계수를 구하고 각 계수에 UMP (Uniform Most Powerful) 테스트를 적용한 후 한 주기 동안 벗어난 계수들의 수의 총합을 이용하여 공정의 이상을 파악하였다^[5,6]. Jin and Shi (2001)는 이상 신호를 모으고 웨이블릿 변환 후, 특정 웨이블릿 계수를 feature로 하여 공정의 이상을 파악하는 방법을 제안하였다^[1]. Jeong 등(2006)은 adaptive threshold를 제안하여 threshold 이상의 값을 가지는 웨이블릿 계수들을 통해 주기 신호 데이터의 변화를 탐지 하는 방법을 제안하였다^[7]. Zhou 등(2006)은 Haar 웨이블릿을 이용하여 웨이블릿 계수를 구하고 특정 수준 이하의 계수들을 분석에 사용하였다^[8]. 선택된 계수에 T^2 차트를 적용하여 이상을 탐지하고, 이상이 탐지됐을 경우 각 계수별 SPC 차트를 통해 이상 상황을 분석하였다.

1.2.3 웨이블릿 계수 선택

웨이블릿을 이용한 이와 같은 연구들은 일반적으로 효과적 이상 탐지를 위해 분석에 사용할 웨이블릿 계수를 선택하고, 선택된 웨이블릿 계수들로 공정의 이상을 파악하였다. 이러한 방법들은 선택된 계수에 따라 이상 탐지

에 대한 성능이 결정되므로 분석에 사용할 웨이블릿 계수를 결정하는 것은 매우 중요하다.

기존 연구들의 웨이블릿 계수 선택에 관한 부분을 살펴보면, Jin and Shi(2001)는 Donoho(1995)^[9]가 제시한 soft threshold를 이용하여, 웨이블릿 계수의 절대값이 threshold를 넘는 계수들을 이상 탐지를 위한 계수로 결정하였다^[1]. 이 방법은 주기 안에서 변동폭이 큰 특성을 가지는 구간에서의 이상 탐지에는 효과적이다. 하지만 변동폭이 좁은 구간에서는 공정에서 관리해야 할 중요한 구간임에도 이 구간에 해당하는 웨이블릿 계수가 선택되지 않거나 정보가 통합된 낮은 수준의 계수가 선택되기 때문에 탐지 못하는 이상이 발생할 수 있다.

Jeong 등(2006)은 웨이블릿 계수 중에서 해당 웨이블릿의 평균과 차이가 큰 계수를 선택하여 T² 차트에 적용하였다^[7]. 이 방법은 각 주기별로 동일한 계수가 선택되지 않을 수 있기 때문에 관리한계선으로 근사값이 사용된다. 따라서 이 방법은 이상 탐지의 신뢰성이 떨어진다.

Zhou 등(2006)은 원래 신호와 각 수준별 scale 계수의 approximation간의 차이를 계산하고, 그 차이가 처음으로 일정 기준 이하로 내려가는 수준을 기준으로 그 수준 이하의 웨이블릿 계수들을 선택하였다^[8]. 이 방법은 주기 신호 데이터의 크기에 비해 잡음이 클 경우 높은 수준(fine level) 계수들까지 선택되는 경향이 있다. 따라서 분석에 필요한 웨이블릿 계수들의 수가 늘어나고, 잡음의 영역이 포함되므로 이상 탐지 성능이 저하된다.

이상의 연구들은 잡음의 크기가 크거나 변동이 적은 특성을 가지는 주기 신호 데이터에는 효과적인 웨이블릿 계수가 선택되지 않는다. 본 연구에서는 잡음의 영향을 줄이고, 주기의 전 구간 걸쳐 발생하는 다양한 이상 상황을 효과적으로 탐지할 수 있는 웨이블릿 계수를 이용한 이상 탐지 방법을 제시하고자 한다. 이를 위해 웨이블릿 잡음제거 방법(wavelet denoise method)과 Haar 웨이블릿 계수의 분산 차이를 이용한 중요 계수 선택 알고리즘을 제안하고 시뮬레이션을 통해 효용성을 확인하고자 한다. 본 연구에서 제시한 방법은 잡음에 영향을 최소화하는 방법이기 때문에 이상 탐지 성능이 기존 방법들보다 우수하며, 신호를 반영하는 중요 계수들만을 사용하기 때문에 자료를 크게 축약할 수 있어서 주기 신호를 가지는 공정의 이상탐지에 효과적인 방법이다.

1.3 논문의 구성

본 논문의 구성은 다음과 같다. 2장에서는 웨이블릿 변환을 소개하고 Haar 웨이블릿에 대하여 설명한다. 3장에

서는 웨이블릿을 이용한 주기 신호 데이터의 이상 탐지 방법 및 제안하는 계수 선택 알고리즘에 대하여 설명하고, 4장에서는 시뮬레이션을 통하여 제안한 알고리즘의 성능을 확인한다. 마지막 5장에서는 본 연구의 결론과 추후 연구방안을 제시한다.

2. 웨이블릿 변환(Wavelet Transform)

2.1 웨이블릿 변환

웨이블릿 변환은 유한한 길이의 기저 함수(basis function)를 사용하여 원래 신호를 표현하는 방법이다. 웨이블릿의 기저함수는 scale 함수 혹은 father 웨이블릿이라 불리는 ϕ 와 detail 함수 혹은 mother 웨이블릿으로 불리는 ψ 로 구성되고 다음과 같이 표현된다^[10].

$$\phi_{j,k}(x) = 2^{j/2}\phi(2^jx - k)$$

$$\psi_{j,k}(x) = 2^{j/2}\psi(2^jx - k)$$

$$\int \phi(x)dx = 1, \int \psi(x)dx = 0$$

여기서 j는 척도모수(scaling parameter)이고 k는 평행 이동모수(translation index)를 의미한다. j가 커질수록 기저함수의 폭은 좁아지고 높이는 높아진다. 또한 k가 커짐에 따라 함수는 우측으로 평행이동하게 된다. 원래 신호는 식 (1)과 같이 L²(R)(힐버트공간)에서 기저 함수의 선형조합으로 표현된다.

$$f(x) = \sum_k c_{j,k}\phi_{j,k}(x) + \sum_j \sum_k d_{j,k}\psi_{j,k}(x) \quad (1)$$

여기서 $c_{j,k}$ 는 j, k에서의 scale 계수, $d_{j,k}$ 는 j, k에서의 detail 계수이다. 그리고 scale 계수와 detail 계수는 피라미드(pyramid) 알고리즘을 통해 다음과 같이 계산된다^[11].

$$c_{j,k} = \langle f(x), \phi_{j,k}(x) \rangle = \sum_n h_{n-2k}c_{j+1,n} \quad (2)$$

$$d_{j,k} = \langle f(x), \psi_{j,k}(x) \rangle = \sum_n (-1)^n h_{-n+2k+1}c_{j+1,n} \quad (3)$$

여기서 h는 기저함수의 변환 계수이고 n은 계수의 시점 인덱스이다. 웨이블릿 변환은 원 신호를 가장 높은 수준의 scale 계수로 놓고 식 (2)와 (3)을 이용하여 하위 수준에 해당하는 scale 계수와 detail 계수를 구한다. 웨이블릿 변환은 변환 수준 별 scale 계수와 detail 계수를 통해 신호를 분석할 수 있기 때문에 다중해상도 분석이 가능하

다. 변환 수준이 높은 계수들은 미시적 관점에서 신호의 세부 영역을 자세히 표현하고, 변환 수준이 낮은 계수들은 거시적 관점에서 전반적인 신호의 변화 형태를 나타낸다.

scale 계수 $c_{j,k}$ 와 detail 계수 $d_{j,k}$ 는 식 (1)을 통해 원 신호를 재구성(reconstruction) 할 수 있다. 그중 0에 가까운 detail 계수는 신호 복원 시에 영향을 거의 주지 못한다. 즉, 이 계수들을 사용하지 않고 원 신호를 재구성해도 원 신호와 크게 차이가 없다. 따라서 축소된 개수의 웨이블릿 계수들로 원래 신호를 표현할 수 있으므로 웨이블릿 변환은 데이터의 크기를 축소할 수 있는 특징이 있다.

2.2 Haar 웨이블릿

본 연구에는 계수 선택을 위한 웨이블릿 변환의 기저 함수로 Haar 웨이블릿을 사용하였다. Haar 웨이블릿은 계산이 매우 빠르고 간단하며 시간 정보에 대한 해석이 용이하다. 또한 평균 변화에 뛰어난 성능을 보이는 특징 때문에 가장 보편적으로 사용되고 있는 기저 함수이다^[5-8]. Haar 웨이블릿은 그림 2와 같은 형태를 가지는 함수이고 다음과 같이 정의 된다.

$$\phi(x) = \begin{cases} 1, & 0 \leq x < 1 \\ 0, & \text{otherwise} \end{cases}$$

$$\psi(x) = \begin{cases} 1, & 0 \leq x < \frac{1}{2} \\ -1, & \frac{1}{2} \leq x < 1 \\ 0, & \text{otherwise} \end{cases}$$

주기 신호 데이터를 2^p 개의 길이를 가진 데이터라 한다면 $X = [x_1, x_2, \dots, x_{2^p}]^T$ 라 표현할 수 있다. Haar 웨이블릿의 변환 계수 h 는 $[\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}]$ 이고 X_h 를 c_{p+1} 로 놓으면 식 (2)와 (3)을 통하여 수준별 웨이블릿 계수의 계산이 가능하다. 최대 수준(coarsest level)까지 변환을 통해 얻어진 Haar 웨이블릿 계수를 c 라 하면 c 는 다음과 같이 표현된다.

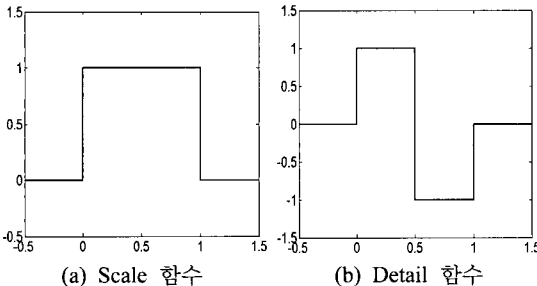


그림 2. Haar 웨이블릿 함수

$$c = [c_{1,1}, d_{1,1}, d_{2,1}, d_{2,2}, \dots, d_{p,2^{p-1}}]^T \quad (4)$$

Haar 웨이블릿 계수들은 식 (5), (6)으로도 계산이 가능하다.

$$c_{j,k} = 2^{\frac{p-j+1}{2}} \bar{f} \{ (k-1)2^{p-j+1} + 1, k2^{p-j+1} \} \quad (5)$$

$$d_{j,k} = 2^{\frac{p-j}{2}} \left[f \left\{ (k-1)2^{p-j} + 1, (k-\frac{1}{2})2^{p-j} \right\} - \bar{f} \left\{ (k-\frac{1}{2})2^{p-j} + 1, k2^{p-j} \right\} \right] \quad (6)$$

여기서, $\bar{f}(i, j) = (1/(j-i+1)) \sum_{k=i}^j x_k$ 이다. 식 (5), (6)

으로부터 scale 계수는 해당 구간의 평균에 비례하고 detail 계수는 이전 구간들의 평균의 차이에 비례함을 알 수 있다. 또한 detail 계수의 부호는 신호의 변동 방향을 의미하고, 계수의 크기는 변동폭을 의미함을 볼 수 있다. 따라서 Haar 웨이블릿 계수는 구간 별 데이터의 평균과 변동에 대한 정보를 제공하므로 다양한 정보를 포함하는 주기 신호 데이터 분석에 적합하다.

3. 웨이블릿을 이용한 이상 탐지 알고리즘

본 연구에서는 Haar 웨이블릿 변환을 통해 계산된 웨이블릿 계수 중에서 잡음에 영향을 받는 계수들을 제거하고 신호의 전 구간의 형태를 잘 반영할 수 있는 계수들을 선택하여 주기 신호 데이터의 이상 탐지 성능을 향상시키고자 한다.

Haar 웨이블릿 계수는 식 (4)와 같이 가장 낮은 수준의 scale 계수와 모든 변환 수준의 detail 계수로 구성 된다. 이 계수들 중 높은 수준의 detail 계수들은 작은 구간에서의 변동을 나타내므로 잡음에 민감하게 변하는 특징이 있

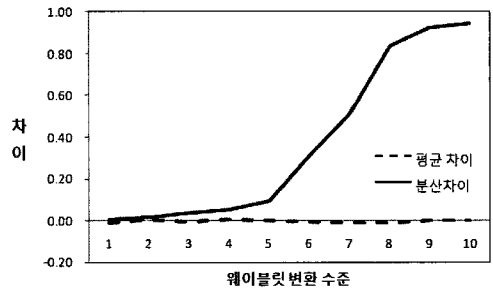


그림 3. 변환 수준별 웨이블릿 계수의 평균과 분산 차이

고, 낮은 수준의 detail 계수들은 큰 구간의 변동을 나타내므로 비교적 잡음의 영향을 받지 않는다. 따라서 원래 신호에서 잡음을 제거하고 나면 높은 수준의 계수들의 값에 변화가 생긴다. 이와 같은 사실을 사용하여 잡음이 제거된 신호와 원 신호의 웨이블릿 계수를 비교하면 잡음에 반응하는 계수들을 파악할 수 있다.

그림 3은 잡음이 제거된 신호와 원 신호의 변환 수준에 따른 detail 계수들 간의 평균과 분산 차이를 나타낸다. 두 계수들 간의 평균은 거의 변화가 없으나 분산은 변환 수준에 따라 차이가 발생함을 볼 수 있다. 높은 변환 수준에서는 분산의 차이가 크고, 변환 수준이 낮아질수록 분산 차이가 크게 감소함을 볼 수 있다. 즉 변환 수준이 낮을수록 잡음의 영향이 줄어드는 것을 볼 수 있다. 따라서 본 연구에서는 웨이블릿 계수의 분산 차이를 이용하여 잡음의 영향을 최소화하면서 주기 신호 형태를 잘 반영하는 웨이블릿 계수를 선택하는 방법을 제안한다.

본 연구에서 제안하는 알고리즘은 다음 세 단계로 구성된다. 첫째 웨이블릿 잡음 제거를 통해 잡음이 제거된 신호를 얻는 단계, 둘째 잡음이 제거된 신호와 원 신호의 Haar 웨이블릿 계수 비교를 통해 주기 신호 데이터의 형태를 반영하는 Haar 웨이블릿 계수를 선택하는 단계, 마지막으로 선택된 계수를 이용하여 SPC 차트를 구축하는 세단계로 구성된다.

3.1 웨이블릿을 이용한 잡음 제거 단계

본 연구에서는 원 신호의 잡음을 제거하기 위하여 웨이블릿 잡음 제거 방법을 사용한다. 효과적인 잡음 제거를 위해 Donoho and Johnstone^[12]이 제시한 soft thresholding 방법을 토대로 잡음 제거 절차를 진행한다. 방법은 다음과 같다.

- i. 웨이블릿 변환에 사용할 웨이블릿 기저함수의 종류와 변환 수준을 선택한다.
- ii. 주기 신호 데이터에 웨이블릿 변환을 적용하여 웨이블릿 계수를 구한다.
- iii. 식 (7)을 이용하여 λ (threshold)값을 구한다.

$$\lambda = \text{median}(|(d^1)|) / 0.6745 \sqrt{2 \log N} \quad (7)$$

여기서 N은 전체 신호의 길이이고 d^1 은 가장 높은 수준의 detail 계수이다.

- iv. 웨이블릿 계수에 thresholding을 적용한다.

$$\hat{d}_{j,k} = \begin{cases} \text{sgn}(d_{j,k})(|d_{j,k}| - \lambda) & \text{if } |d_{j,k}| \geq \lambda \\ 0 & \text{if } |d_{j,k}| < \lambda \end{cases} \quad (8)$$

- v. Thresholding 된 웨이블릿 계수로 식 (1)을 이용하여 신호를 복원한다.

잡음 제거 시에 사용되는 웨이블릿 기저함수와 변환 수준(decomposition level)은 잡음 제거 성능을 좌우하는 중요한 요소이다. 데이터의 형태나 잡음의 크기에 따라 해당 신호를 잘 나타낼 수 있는 웨이블릿 함수와 변환 수준은 다르기 때문이다. 보통 계단 형태의 데이터는 Haar 웨이블릿이, 곡선 형태의 데이터는 Daubechies나 Coiflet 웨이블릿이 잡음 제거에 좋은 성능을 보이는 것으로 알려져 있다^[13,14].

본 연구에서는 주어진 신호의 잡음을 가장 잘 제거할 수 있는 웨이블릿 기저함수와 변환 수준을 찾기 위하여 다양한 기저함수와 변환 수준에 대하여 MSE(Mean Squared Error)를 비교한다. 비교를 위한 MSE는 식 (9)와 같이 정의된다.

$$MSE = \frac{1}{N} \sum_{i=1}^N \sum_{t=1}^2 (\hat{X}_i(t) - T(t))^2 \quad (9)$$

여기서 i는 샘플 인덱스이고 t는 시점 인덱스이다. \hat{X}_i 는 soft threshold를 통해 잡음이 제거된 i번째 샘플 데이터이고, N은 샘플의 개수, T는 정상 상태를 나타내는 이상적인 데이터이다. 실제 적용 시에는 이상적인 데이터를 얻을 수 없기 때문에 모여진 데이터의 평균값을 T(t)로 사용한다. MSE 값이 작다는 것은 이상적인 데이터와 차이가 작다는 것, 즉 잡음이 가장 잘 제거된 신호를 의미한다. 따라서 적용한 함수와 수준 중에서 MSE 값이 최소가 되는 웨이블릿 기저함수와 변환 수준을 선택하여 가장 잡음이 잘 제거된 신호를 얻는다.

3.2 Haar 웨이블릿 계수 선택 단계

주기 신호 데이터의 형태를 잘 반영하는 웨이블릿 계수를 얻기 위하여 원 신호와 앞장에서 구한 잡음이 제거된 신호에 각각 Haar 웨이블릿 변환을 시행하고 웨이블릿 계수를 구한다. Haar 웨이블릿 계수들의 분산을 계산하고 계수들 간의 분산 차이를 비교한다. 본 연구에서는 두 계수간의 분산 차이를 측정하기 위한 척도 D_j 을 식 (10)으로 정의하였다.

$$D_l = \frac{1}{N} \sum_{i=1}^N \frac{(\theta_{i,l} - \hat{\theta}_{i,l})^2}{\theta_{i,l}^2} \quad (10)$$

여기서 l 은 Haar 웨이블릿 계수의 인덱스를 의미하다. θ 는 원 데이터의 Haar 웨이블릿 계수의 분산이고, $\hat{\theta}$ 는 잡음이 제거된 데이터의 Haar 웨이블릿 계수의 분산이다. D_l 이 일정 수준(γ) 이하이면 잡음의 영향이 없고 신호의 형태를 잘 반영하는 계수로 판단하여 분석에 필요한 중요 계수 \tilde{c} 로 선택한다.

3.3 SPC 차트 구축 단계

이번 단계는 앞장에서 선택한 Haar 웨이블릿 계수를 이용하여 주기 신호 데이터의 이상을 탐지하기 위한 T^2 차트를 구축하는 단계이다. 효율적인 T^2 차트를 만들기 위하여 Phase I, Phase II 단계를 통하여 관리한계선을 결정한다^[15].

$$T_i^2 = (\tilde{c}_i - \bar{\tilde{c}})^T S^{-1} (\tilde{c}_i - \bar{\tilde{c}}) \quad (11)$$

$$S = \frac{1}{N-1} \sum_{i=1}^N (\tilde{c}_i - \bar{\tilde{c}})(\tilde{c}_i - \bar{\tilde{c}})^T \quad (12)$$

여기서 $\bar{\tilde{c}}$ 는 \tilde{c}_i 의 평균이고 S 는 \tilde{c}_i 의 공분산이다. Phase I의 관리한계선은 Type I 에러를 α 라 하면

$$UCL_1 = \frac{(N-1)^2}{N} \beta_{\alpha, N_c, (N-N_c-1)/2} \quad (13)$$

로 정해진다. 여기서 N_c 는 \tilde{c}_i 의 길이이고, $\beta_{\alpha, N_c, (N-N_c-1)/2}$ 는 N_c 과 $(N-N_c-1)/2$ 를 모수로 하는 베타 분포의 $1-\alpha$ 값을 의미한다. 모인 데이터에 Phase I의 관리한계선을 적용하여 이상 데이터를 제거한 후 남은 데이터로 수정된 관리한계선을 정한다. 이와 같은 과정을 반복하여 이상 데이터에 대한 분석 및 처리가 완료되면 Phase II 단계를 진행한다.

Phase II에서 Phase I에서 남겨진 데이터들을 이용하여 $\bar{\tilde{c}}$ 와 S^{-1} 를 구하고 식 (14)를 통해 새로운 관리한계선을 계산한다.

$$UCL_2 = \frac{N_c(N-1)(N+1)}{N^2 - N_c N} F_{\alpha, N_c, N-N_c} \quad (14)$$

여기서 $F_{\alpha, N_c, N-N_c}$ 는 N_c 과 $N-N_c$ 를 모수로 하는 F 분포의 $1-\alpha$ 값이다.

주기 신호 데이터의 이상 탐지는 Haar 웨이블릿 변환을 통하여 웨이블릿 계수를 구하고, 선택된 계수와 동일한 인덱스를 가진 계수를 사용하여 Phase II에서 정한 UCL_2 를 통해 이루어진다.

4. 실험 및 결과 검증

4.1 시뮬레이션 데이터

시뮬레이션에 사용한 주기 신호 데이터는 Johnson S_u 분포 JSU(0, 20, 2, 6)과 JSU(-2, 5, 5, 2)와 JSU(-3, 6, 6, 4) 분포 3개가 합해진 1024의 길이를 가지는 신호로 그림 4와 같은 형태를 지니고 있다. 이 데이터는 구간의 모양에 따라 100에서 300의 구간, 300에서 800 구간, 800에서 1000구간으로 분류할 수 있다. 정상신호의 데이터는 평균이 0이고 분산이 1인 정규분포를 가지는 노이즈를 더하여 200개를 생성하였다.

4.2 중요 계수 선택 및 SPC 차트 구축

본 시뮬레이션에 사용된 데이터는 곡선 형태의 데이터이므로 잡음 제거를 위한 웨이블릿 기저함수로 Daubechies, Coiflet 함수를 사용하여 잡음제거를 실시하였다. Daubechies 2~20와 Coiflet 1~4를 변환 수준별로 soft thresholding 방법을 통하여 MSE를 계산하였다. 결과는 표 1과 같다. MSE 비교를 통하여 Daubechies 14의 6 수준으로 잡음 제거 하는 것이 가장 효과적인 방법으로 선택되었다.

γ 를 0.05로 정하고 Daubechies 14의 6 수준으로 잡음을 제거한 데이터와 원 데이터를 기준으로 D_l 을 계산하면 그림 5와 같다. 그림 5에서 l 이 32 이상인 계수들은 D_l 값이 0.2 이상이므로 표현하지 않았다. 분석에 사용할 계수로 D_l 이 0.05 이하의 값을 갖는 16개의 Haar 웨이블릿 계수가 선택되었다. 따라서 본 실험에서는 제안된 알고리즘

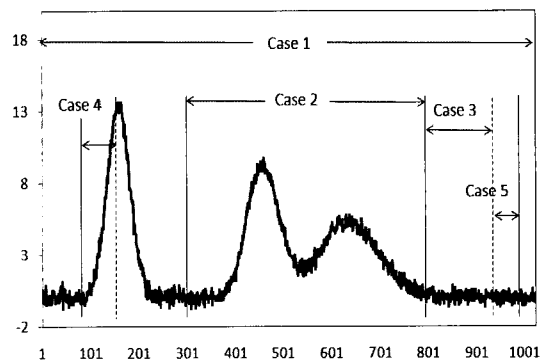


그림 4. 시뮬레이션 주기 신호 데이터 형태

표 1. 웨이블릿 함수와 수준별 MSE

웨이블릿 함수 종류	수준	level 1	level 2	level 3	level 4	level 5	level 6	level 7	level 8	level 9	level 10
	Daubechies 2		18.26	18.09	17.86	17.33	16.47	15.1	13.55	13.19	16.42
Daubechies 4		62.68	23.55	22.83	19.5	11.66	9.92	9.21	11.54	16.08	22.63
Daubechies 6		98.4	68.2	66.91	39.04	14.1	8.69	8.57	11.51	16.05	22.62
Daubechies 8		116.27	93.9	85.32	45.76	25.04	8.08	8.44	11.49	16.01	22.61
Daubechies 10		121.6	108.09	92.18	55.27	14.77	7.92	8.39	11.45	16	22.59
Daubechies 12		122.79	115.65	94.04	75.25	31.5	7.72	8.37	11.45	16.02	22.59
Daubechies 14		124.98	121.64	96.09	92.25	41.47	7.2	8.35	11.47	16.04	22.59
Daubechies 16		125.21	123.84	96.35	95.89	38.39	8.11	8.34	11.47	16.02	22.6
Daubechies 18		124.88	124.4	98.76	94.27	30.45	9.03	8.33	11.44	15.99	22.6
Daubechies 20		125.04	124.89	104.33	94.63	30.65	8.29	8.32	11.42	15.97	22.6
Coiflet 1		107.88	107.87	76.39	74.77	19.14	10.09	9.11	11.51	16.02	22.59
Coiflet 2		124.83	124.83	124.81	100.01	86.26	8.4	8.28	11.38	15.96	22.57
Coiflet 3		125.18	125.18	125.18	124.65	86.87	77.65	8.13	11.29	15.94	22.55
Coiflet 4		125.26	125.26	125.26	125.26	106.81	86.28	9.02	11.25	15.86	22.53

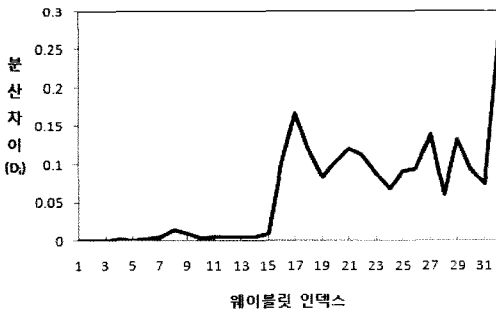


그림 5. 웨이블릿 계수 분산 차이

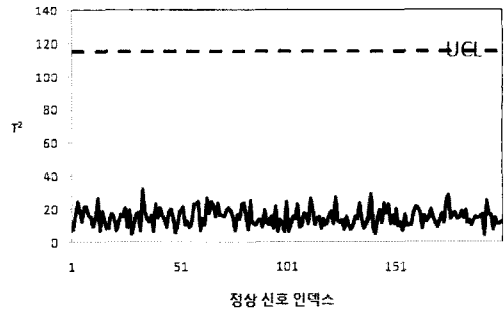


그림 6. Phase I T2 차트

을 통해 선택된 16개의 Haar 웨이블릿 계수로 1024의 시점 갖는 주기 신호 데이터의 이상을 탐지하였다.

선택된 계수들을 사용하여 α 수준 0.05로 Phase I에 적용한 결과, UCL_1 은 117.4로 계산되고 정상 신호에 대한 T^2 차트는 그림 6과 같이 계산된다. 모두 정상 상태로 파악되므로 phase II의 관리한계선($UCL_2 = 29.5$)을 사용하여 이상 상황에 대한 평가를 수행하였다.

4.3 성능 평가

본 알고리즘의 성능 평가를 위해서 다음 세 가지 방법과 비교하였다. 방법 (1)은 전통적인 다변량 통계분석 방법이다. 각 시점을 변수로 보고 χ^2 차트를 통하여 이상을 탐지 하는 방법이다. 1024개의 시점을 각각의 변수로 보

고 χ^2 를 계산한다. 방법 (2)는 Jin and Shi가 제안한 threshold를 넘는 계수들을 중요 계수를 사용하는 방법이다^[1]. 시뮬레이션 데이터에 적용한 결과, 웨이블릿 계수의 절대값이 threshold 값 2.44를 초과하는 32개의 계수가 선택되었다. 방법 (3)은 Zhou 등이 제안한 Q index를 통해 중요 계수를 결정하는 방법이다^[8]. Q값을 5%로 하였을 때 시뮬레이션 데이터에 적용한 결과, 7 수준이 선택되었고 7 수준 이하의 128개의 웨이블릿 계수가 선택되었다. 이 세 가지 방법과 본 연구에서 제시한 알고리즘을 통해 얻은 16개의 웨이블릿 계수를 사용하는 방법(4)로 하여 다양한 이상 상황에 대하여 검출 성능을 비교하였다. 네 방법 모두 α 수준은 0.05로 동일하게 적용하였다.

주기 신호 데이터는 한 주기 안에 다양한 형태의 패턴

을 가지고 있다. 이러한 주기 신호 데이터 안에서 공정의 이상은 다양한 구간에서 발생할 수 있다. 따라서 본 시뮬레이션에서는 다양한 상황에 대한 성능 평가를 위해 그림 3과 같이 패턴의 형태와 구간의 크기에 따라 발생할 수 있는 이상 상황을 5가지로 구분하였다. 1) 전체 구간, 2) 변화가 존재하는 넓은 구간, 3) 평평한 넓은 구간, 4) 변화가 존재하는 좁은 구간, 5) 평평한 좁은 구간으로 분류하고, 각 구간의 평균을 0.1 단위로 0에서 2까지 변화시켜 평균 변화에 대한 검출 성능을 비교하였다. 성능 평가는 각 이상 상황에 대해 10000개의 데이터를 생성한 후 각 경우에 적용하여 이상 상황을 탐지한 수의 비율로 평가하였다.

Case 1. 전 구간의 평균 변화

이 경우는 공정 계측 장비의 오류와 전체 신호의 특성 변화 등으로 생기는 이상 상황으로 전 구간에 걸쳐 평균의 변화가 발생한 경우이다. 다양한 크기의 평균 변화를 네 가지 방법에 적용해 본 결과는 그림 7과 같다. 본 연구에서 제안한 방법은 0.1에서 0.5까지의 평균 변화에 대해 다른 방법들보다 우수한 검출 성능을 보였다. 0.5 이상의 평균 변화에는 네 가지 방법 모두 동일하게 이상을 탐지하

였다.

Case 2. 변화가 존재하는 넓은 구간의 평균 변화

실제 공정에서는 주기 신호 안에 다양한 공정이 진행되므로 전체 시간의 변화보다 부분적인 변화가 빈번히 발생한다. 주기 신호 데이터는 대부분 복잡한 형태를 가지고 있다. 이번 경우에는 변화가 크면서 넓은 구간에서의 성능을 평가하기 위해 시점 300에서 800까지의 구간에 평균 변화를 주어 평가하였고 결과는 그림 8로 표현된다. 본 연구에서 제안한 방법이 0.1에서 0.7까지의 평균 변화에 대해 다른 방법들보다 우수한 검출 성능을 보였다.

Case 3. 평평하고 넓은 구간의 평균 변화

공정 특성에 따라 평평한 형태를 유지하는 공정 구간이 있다. 이러한 영역에서 발생하는 이상 상황의 성능 평가를 위해 시점 800에서 950까지 구간의 평균을 변화시켜 검출 능력을 평가하였다. 결과는 그림 9와 같다. 본 연구에서 제안한 방법이 0.1에서 1.2까지의 평균 변화에 대해 다른 방법들보다 우수한 검출 성능을 보였다.

Case 4. 변화가 존재하고 좁은 구간의 평균 변화

패턴 변화가 크면서 좁은 구간에서의 성능 평가를 위해 가장 급격하게 변하는 시점 100에서 150까지의 구간

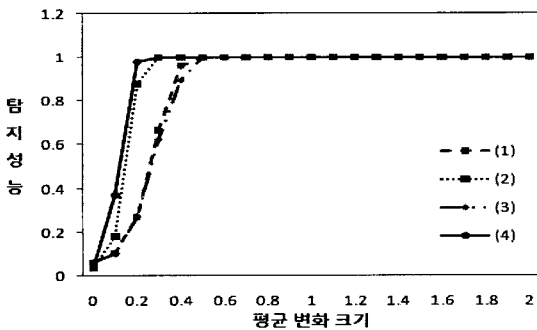


그림 7. 전 구간 평균 변화 성능 비교

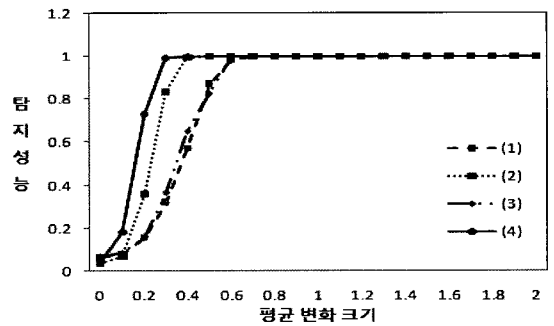


그림 8. 변화가 존재하는 넓은 구간의 평균 변화 성능 비교

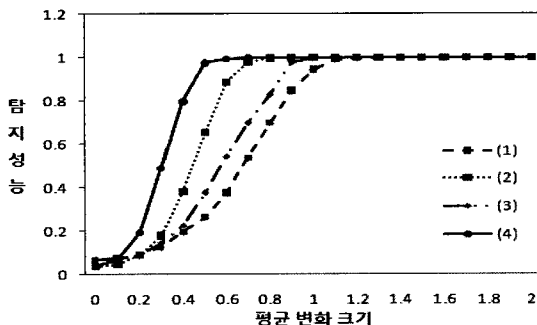


그림 9. 평평하고 넓은 구간의 평균 변화 성능 비교

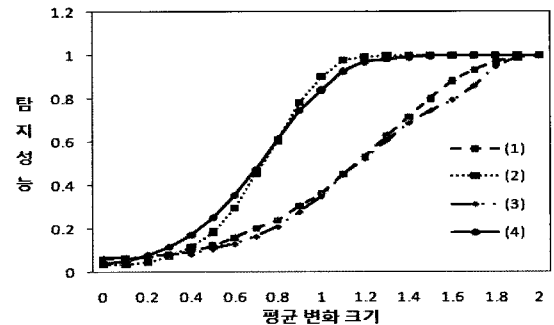


그림 10. 변화가 존재하는 좁은 구간의 평균 변화 성능 비교

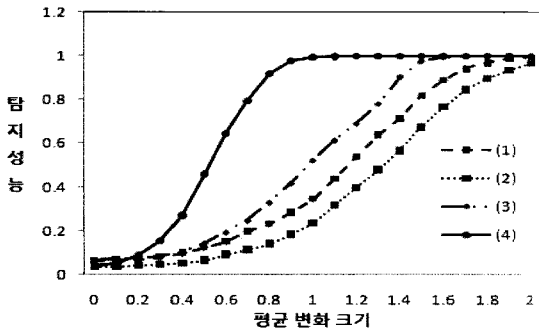


그림 11. 평평한 좁은 구간 평균 변화의 성능 비교

에 평균 변화를 주어 검출 능력을 평가하였고 결과는 그림 10과 같이 나타났다. 0.7 이전의 평균 변화에는 본 알고리즘이 우수한 성능을 보였고 0.8 이상의 평균 변화에 대해서는 대안 2에서 제안한 알고리즘의 성능이 우수하였다. 대안 2의 알고리즘은 변화가 큰 계수들을 선택하므로 변화가 큰 구간에서 보다 우수한 성능을 보임을 볼 수 있었다.

Case 5. 평평한 좁은 구간의 평균 변화

평평하면서 좁은 구간의 성능은 시점 900에서 950까지의 구간에 평균 변화를 주어 검출 능력을 비교하였고 결과는 그림 11과 같다. 본 알고리즘이 다양한 평균 변화에 대해서 다른 방법들 보다 월등한 성능을 보였다. 반면에 대안 2의 방법은 가장 낮은 성능으로 보였다. 대안 2의 방법은 평평한 구간의 계수들은 선택되지 않기 때문에 효과적인 검출이 어렵다.

시뮬레이션 결과를 정리하면 본 연구에서 제안한 방법이 동일한 신뢰 수준에서 Case 4의 일부 구간을 제외한 모든 경우에서 가장 우수한 이상 탐지 능력을 보였다. 구간이 좁아질수록 더 좋은 성능을 보였고, 특히 평평하고 변화가 작은 구간의 이상 탐지에 탁월한 성능을 보였다. 또한, 본 알고리즘을 통해 선택된 계수의 수는 16개로 네 가지 방법 중에서 가장 작았다. 따라서 본 알고리즘은 이상 탐지의 성능뿐만 아니라 데이터 축소 효과 또한 뛰어난 것을 볼 수 있었다.

5. 결 론

본 연구에서는 주기 신호 데이터의 이상을 탐지하기 위해 웨이블릿 잡음 제거와 Haar 웨이블릿 계수의 분산 차이를 이용하여 데이터의 형태를 반영하는 중요 계수를 선택하고, 선택된 계수로 SPC 차트를 구축하였다. 잡음 제거 시에 다양한 웨이블릿 기저함수와 변환 수준을 적용하고 이들의 MSE 비교를 통해 신호의 형태에 적합한 잡

음제거를 시행하였다. 제안된 방법은 시뮬레이션을 통하여 동일 신뢰 수준에서 기존방법들보다 다양한 구간, 다양한 크기의 평균 변화에 대하여 뛰어난 이상 탐지 능력을 가짐을 확인하였다. 또한 잡음의 영향을 최소화하는 웨이블릿 계수 선택을 통해 데이터의 길이를 급격하게 줄이는 효과도 확인하였다.

추후 연구 과제로는 실제 공정에 적용하여 실제로 발생하는 다양한 이상 상황에 대한 효과를 검증해보는 연구가 필요하다. 또한 이상을 탐지한 후 문제가 되는 웨이블릿 계수를 통해 이상의 형태를 파악하고 원인을 진단할 수 있는 연구가 필요하다.

참 고 문 헌

1. Jin, J. and Shi, J. (2001) "Automatic feature extraction of waveform signals for in-process diagnostic performance improvement," *Journal of Intelligent Manufacturing*, Vol. 12, No. 3, pp. 267-268.
2. 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. 22, pp. 360-369.
3. Zhou, S. and Jin, J. (2005), "Automatic feature selection for unsupervised clustering of cycle-based signals in manufacturing processes", *IIE Transactions*, Vol. 37, No. 6, pp. 569-584.
4. Bakshi, B. R. (1998) "Multiscale PCA with application to multivariate statistical process monitoring", *AIChE Journal*, Vol. 44, No. 7, pp. 1596-1610.
5. Koh, C. K. H., Shi, J., Williams, W. J. and Ni, J. (1999), "Multiple fault detection and isolation using the haar transform, Part 1: Theory", *Journal of Manufacturing Science and Engineering*, Vol. 121, No. 2, pp. 290-294.
6. Koh, C. K. H., Shi, J., Williams, W. J. and Ni, J. (1999) "Multiple fault detection and isolation using the haar transform, Part 2: Application to the stamping Process", *Journal of Manufacturing Science and Engineering*, Vol. 121, No. 2, pp. 295-299.
7. Jeong, M. K., Lu, J. C. and Wang, N. (2006), "Wavelet-based SPC procedure for complicated functional data". *International Journal of Production Research*, Vol. 44, No. 4, pp. 729-744.
8. Zhou, S., Sun, B. and Shi, J. (2006), "An SPC monitoring system for cycle-based waveform signals using haar transform", *IEEE Transactions on Automation Science and Engineering*, Vol. 3, No. 1, pp. 60-72.
9. Donoho, D. L. (1995), "Denosing by soft thresholding",

- IEEE Transactions on Information Theory*, Vol. 41, No. 3, pp. 613-627.
10. Burrus, C. S., Gopinath, R. A. and Guo, H. (1998), *Introduction to Wavelets and Wavelet Transforms*, A Primer. Upper Saddle River, NJ: Prentice-Hall.
 11. Burt, P. J. and Adelson, E. H. (1983), "The Laplacian pyramid as a compact image code", *IEEE Transactions on Communications*, Vol. 31, No. 4, pp. 532-540.
 12. Donoho, D. L. and Johnstone, I. M. (1994), "Ideal spatial adaptation by wavelet shrinkage", *Biometrika*, Vol. 81, No. 3, pp. 425-455.
 13. Ganesan, R., Das, T. K. and Venkataraman, V. (2004), "Wavelet-based multiscale statistical process monitoring: A literature review", *IIE Transactions*, Vol. 36, No. 9, pp. 787-806.
 14. Nounou, M. N. and Bakshi, B. R. (1999), "On-line multi-scale filtering of random and gross errors without process models", *AIChE Journal*, Vol. 45, No. 5, pp. 1041-1058.
 15. Montgomery, D. C. (2001), *Introduction to Statistical Quality Control*, 4th Edition, Hohn Wiley & Sons, New York, NY.



이재현 (zoeforever@korea.ac.kr)

2006 고려대학교 공과대학 산업시스템정보공학과 학사
2006~현재 고려대학교 공과대학 산업시스템정보공학과 석사과정

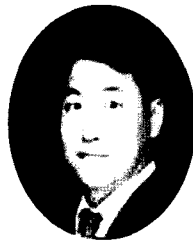
관심분야 : Data Mining, Quality Control, System Optimization



김지현 (jihyunk@korea.ac.kr)

1993 고려대학교 공과대학 산업공학과 학사
1995 고려대학교 공과대학 산업공학과 석사
1995 삼성전자 반도체 총괄 메모리사업부
2004 미국 University of Michigan 산업공학과 박사
2004~2007 고려대학교 정보통신기술연구소 연구교수
2007~현재 고려대학교 반도체기술연구소 연구교수

관심분야 : Advanced Process Control, Quality Control



황지빈 (ji_bin_@hanmail.net)

2006 고려대학교 공과대학 산업시스템정보공학과 학사
2006~현재 고려대학교 공과대학 산업시스템 정보공학과 석사과정

관심분야 : System Optimization, Financial Engineering



김성식 (sungskim@korea.ac.kr)

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

관심분야 : Advanced Process Control, Modeling, System Optimization