

CART 기반의 변수랭킹을 이용한 곱셈형 고장의 고장변수식별

유정원, 김진홍, 이영재, 임길택

한국전자통신연구원 대경권연구센터

{gardenyoo, jinhong, lyj4295, ktl}@etri.re.kr

Faulty variable isolation of multiplicative fault using CART-based variable ranking

Jungwon Yu, Jinhong Kim, Youngjae Lee and Kil-Taek Lim

Electronics and Telecommunications Research Institute

요약

고장변수식별(faulty variable isolation)은 대상공정에서 발생된 고장(fault)과 가장 밀접하게 연관된 공정변수를 식별하는 작업이다. 복잡한 산업공정의 고장변수를 정확하게 식별하는 작업은 운전원과 엔지니어가 고장을 정확하게 진단(diagnosis)할 수 있도록 도움을 준다. 본 논문에서는, classification and regression tree (CART) 기반의 변수랭킹 방법을 곱셈형 고장(multiplicative fault)의 고장변수식별을 위해 적용한 결과를 제시한다. 실험결과는 CART 기반의 변수랭킹 방법이 되물음 효과(smearing effect) 없이, 고장변수를 정확하게 식별할 수 있음을 보여준다.

I. 서론

복잡한 산업공정에서, 적시에 정확하게 발생된 고장(fault)과 연관된 공정변수(process variable)를 식별하는 작업은 안전하고, 효율적이며, 경제적인 공정운전을 위해서 필수적이다. 여기서 ‘고장’이란 공정의 운전 중에 발생하는 비정상적인 이벤트를 의미한다. 고장변수(faulty variable)를 식별하기 위해서 가장 널리 사용되는 방법은 기여도 분석(contribution analysis)이다. 기여도 분석은 고장탐지(fault detection)를 위해서 계산된 탐지지수(detection index)에 대한 각 변수의 기여도 값을 계산한 후, 계산된 기여도 값이 높은 변수를 고장변수로 식별한다. 기여도 분석은 발생 가능한 고장유형(fault type)에 대한 사전지식과 과거 고장데이터 없이, 빠르고 손쉽게 고장변수를 식별할 수 있지만, 되물음 효과(smearing effect)로 인해 공정 운전원이나 엔지니어에게 혼란을 초래할 수 있다. ‘되물음 효과’란, 고장이 탐지된 이후에 수행되는 기여도 분석에서 고장변수의 기여도 값 뿐만 아니라 정상변수의 기여도 값 역시 변화되는 현상이다. 복잡한 고장이 발생했을 경우, 기여도 분석에 의한 고장변수식별은 상당히 애매모호한 진단(diagnosis)결과로 이어질 수 있다.

본 논문에서는, 참고문헌 [1]에서 제안된 classification and regression tree (CART) 기반의 변수랭킹(variable ranking) 방법을 곱셈형 고장(multiplicative fault)의 고장변수식별(faulty variable isolation)을 위해서 적용한 결과를 제시한다. 적용결과는 CART 기반의 변수랭킹 방법이 정상변수로서의 되물음 없이 고장변수를 정확하게 식별할 수 있음을 보여준다.

II. CART 기반의 변수랭킹 방법

참고문헌 [1]에서 제안된 CART 기반의 변수랭킹을 이용한 고장변수식별 방법에서는, 고장샘플(faulty sample)과 정상샘플(normal sample)로 구성된 학습데이터에 CART 알고리즘을 적용하여 이진분류트리(binary classification tree)를 건설한 후, 최종트리(final tree)로부터 각 입력변수의 중요도(variable importance) 값을 추출하여 변수랭킹을 수행한다. 정상 클래스가 할당된 정상샘플은 대상공정이 정상적으로 동작할 때 수집된

공정데이터이고, 고장 클래스가 할당된 고장샘플은 고장탐지기법에서 탐지지수가 문턱값(threshold)을 위배한 시간구간에 해당하는 공정데이터이다. 계산된 중요도 값이 높은 변수들은 원 입력공간(original input space)에서 정상샘플과 고장샘플을 분류하기 위한 중요한 변수들이다. 즉, 클래스를 구분하기 위한 설명력(explanatory power)이 높은 변수들이다.

참고문헌 [1]에는 CART 기반의 변수랭킹을 이용한 고장변수식별을 위해서, 두 가지 접근방법이 사용되었다. 첫 번째 접근방법(‘방법 1’)은 전체 고장구간에 대하여 발생한 고장과 가장 밀접하게 연관된 고장변수를 식별하기 위함이고, 두 번째 접근방법(‘방법 2’)은 time window sliding을 이용하여 고장이 어떻게 과급되는지를 관찰하기 위함이다.

III. 실험결과

모니터링 변수의 공분산 구조(covariance structure)에 이상이 생기는 고장을 ‘곱셈형 고장’이라고 정의한다 [2]. 곱셈형 고장의 경우, 곧바로 시스템실패(system failure)로 이어지지 않을 수도 있지만, 대상 설비의 수명과 밀접한 관련이 있다. 본 논문에서는, 참고문헌 [2]에서 사용된 곱셈형 고장과 관련된 벤치마크(benchmark) 문제에 CART 기반의 변수랭킹방

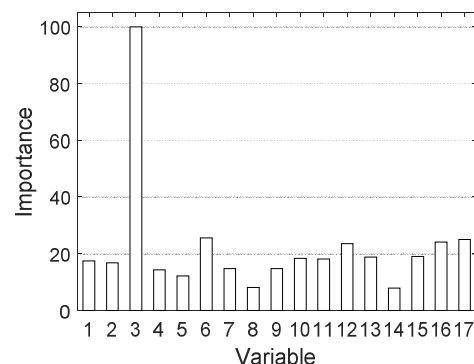


그림 1. ‘방법 1’에 의해서 전체 고장구간($k = 101, \dots, 500$)에서 계산된 변수 중요도 값 [3].

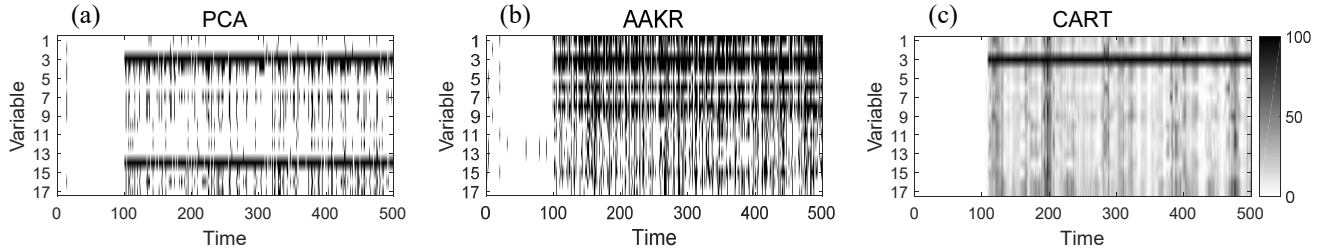


그림 2. 비교방법 및 ‘방법 2’에 의한 고장변수식별 결과: (a) 주성분분석 기반의 기여도 분석; (b) AAKR 기반의 기여도 분석; (c) CART 기반의 변수랭킹을 이용한 고장변수 식별 [3].

법을 적용하여 고장변수식별 작업을 수행한다. 시뮬레이션을 위한 데이터는 아래의 식[2]에 의해서 생성된다.

$$\begin{aligned} \mathbf{y}(k) &= \mathbf{W} \mathbf{x}(k) + \boldsymbol{\xi}(k) \\ \boldsymbol{\theta}(k) &= \boldsymbol{\Psi} \mathbf{y}(k) + \mathbf{b} + \boldsymbol{\eta}(k) \end{aligned} \quad (1)$$

여기서 $\mathbf{x}(k) \in R^3$, $\mathbf{y}(k) \in R^{15}$, $\boldsymbol{\theta}(k) \in R^2$, $\mathbf{W} = rand(15, 3) \in R^{15 \times 3}$, $\boldsymbol{\Psi} = 1 + rand(2, 15) \in R^{2 \times 15}$, $\mathbf{b} = [1055, 1825]^T$, $\mathbf{x} \sim MVN_3(0, diag[4, 1.96, 1.44])$, $\boldsymbol{\xi} \sim MVN_{15}(0, 10^{-16} \mathbf{I}_{15})$, $\boldsymbol{\eta} \sim MVN_2(0, 10^{-8} \mathbf{I}_2)$ 이고, 행렬 $rand(\cdot, \cdot)$ 의 각 성분은 0에서 1 사이의 균등분포를 따른다. 식 (1)로부터, 17개의 공정변수 $[\mathbf{y}^T(k), \boldsymbol{\theta}^T(k)]^T$ 로 구성된 500개의 정상샘플을 생성하여, 다변량데이터 기반의 고장탐지 모델의 학습을 위해 사용하며, 이 정상데이터에는 분류트리를 건설하기 위해 정상클래스가 할당된다. 고장변수식별의 성능을 검증하기 위해서, 고장데이터는 다음과 같은 시나리오에 의해서 생성된다. 먼저 식 (1)로부터 500개의 정상샘플을 생성한 후에, 곱셈형 고장이 발생한 시간구간($k = 101, \dots, 500$)에 대해 식 $\mathbf{y}(k) = \mathbf{F} \mathbf{y}(k)$ 에 의해서 곱셈형 고장효과(fault effect)가 주입된다. 여기서 행렬 $\mathbf{F}$ 은 고장구간에서는 $\mathbf{F}(3, 6) = \mathbf{I}_{15}(3, 6) + 5$ 로 정의되며, 정상구간에서는 $\mathbf{F} = \mathbf{I}_{15}$ 를 만족한다.

먼저 ‘방법 1’을 이용한 전체 고장구간($k = 101, \dots, 500$)에서의 고장변수 식별결과에 대해서 살펴보자. 그림 1[3]은 위에서 설명한 과정에 의해서 생성된 곱셈형 고장데이터에 ‘방법 1’을 적용하여 계산된 변수중요도이다. 그림 1에서 확인할 수 있듯이, 고장변수 y_3 의 중요도 값은 100이며, 나머지 정상변수의 중요도 값은 30보다 작다.

그림 2[3]는 주성분분석(principal component analysis, PCA) 및 auto-associative kernel regression(AAKR)에 기반한 기여도 분석과 ‘방법 2’를 곱셈형 고장데이터에 적용하여 얻어진 고장변수식별 결과를 보여 준다. 주성분분석 기반의 기여도 분석은 고장변수인 y_3 뿐만 아니라, 정상변수인 y_{14} 역시 고장변수로 식별한다. 즉, y_{14} 에 되물음이 발생하였다. AAKR에 기반한 기여도 분석에서는, 되물음 효과가 더 심하게 발생하며, 정상변수 y_1, y_4, y_6, y_8 에서 되물음이 발생함을 관찰할 수 있다. ‘방법 2’의 경우, 고장변수 y_3 의 중요도 값은 항상 100이다. y_3 를 제외한 나머지 정상변수의 중요도 값은 시간이 지남에 따라 변화가 있지만, 그 크기는 고장변수의 중요도보다 항상 작다.

IV. 결 론

CART 기반의 변수랭킹을 이용한 고장변수식별 방법[1]은 원 공간에서 정상샘플과 고장샘플을 구분하기 위해서 설명력이 높은 공정변수를 고장

변수로 식별한다. 즉, 원 공간에서 결정경계(decision boundary)를 설정하는 분류트리에 기반하여 고장변수식별이 수행되므로, 되물음 효과가 발생하지 않는다. 본 연구에서는, CART 기반의 변수랭킹을 이용한 방법을 곱셈형 고장에 적용한 결과를 제시하였다. 실험결과로부터 CART에 기반한 방법이 비교방법인 주성분분석 및 AAKR에 기반한 기여도 분석에 비해서 더 명확하게 곱셈형 고장의 고장변수를 식별할 수 있음을 확인할 수 있다.

ACKNOWLEDGMENT

This work was supported by Electronics and Telecommunications Research Institute (ETRI) grant funded by the Korean government. [21ZD1120, Development of ICT Convergence Technology for Daegu-GyeongBuk Regional Industry]

참 고 문 헌

- [1] J. Yu, J. Jang, J. Yoo, J. H. Park, and S. Kim, “A Fault Isolation Method via Classification and Regression Tree-Based Variable Ranking for Drum-Type Steam Boiler in Thermal Power Plant. Energies, vol. 11, no. 5, pp. 1-19, 2018.
- [2] H. Hao, K. Zhang, S. X. Ding, Z. Chen, and Y. Lei, “A data-driven multiplicative fault diagnosis approach for automation processes,” ISA Trans., vol. 53, no. 5, pp. 1436-1445, Sep. 2014.
- [3] J. Yu, “Data-Driven Fault Detection and Isolation of Complex Industrial Processes Using Hybrid of Multivariate Statistical Techniques,” Ph.D. dissertation, Dept. of Elect. and Comput. Eng., Pusan Nat. Univ., Busan, South Korea, 2018.