



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2018-0039205
(43) 공개일자 2018년04월18일

(51) 국제특허분류(Int. Cl.)
G05B 23/02 (2006.01)

(52) CPC특허분류
G05B 23/0283 (2013.01)
G05B 23/0221 (2013.01)

(21) 출원번호 10-2016-0129702

(22) 출원일자 2016년10월07일
심사청구일자 2016년10월07일

(71) 출원인
고려대학교 산학협력단

서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)

(72) 발명자
백준걸

서울특별시 강동구 고덕로 131, 122동 1503호(암사동, 강동롯데캐슬퍼스트아파트)

장성원

서울특별시 동대문구 왕산로 52 101동 1405호 (용두동, 신설동역대우아이빌아파트)

박승환

경기도 수원시 영통구 영통로90번길 4-27 113동 901호 (망포동, 늘푸른벽산아파트)

(74) 대리인

김동용, 김홍석

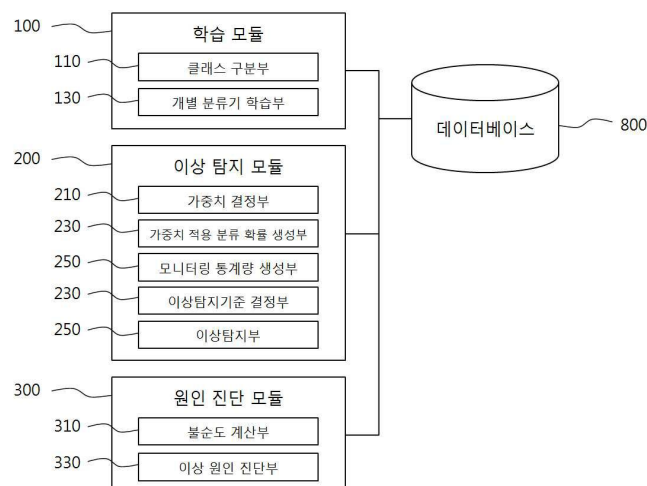
전체 청구항 수 : 총 9 항

(54) 발명의 명칭 개선된 RTC(Real-time contrasts) 기법을 이용한 지능형 이상 진단 방법 및 장치

(57) 요약

학습 모듈, 이상 탐지 모듈, 이상 원인 진단 모듈 및 데이터베이스를 포함하는 개선된 RTC(Real-time contrasts) 기법을 이용한 지능형 이상 진단 장치 및 이를 이용한 개선된 RTC 기법을 이용한 지능형 이상 진단 방법이 개시된다. 개선된 RTC 기법을 이용한 지능형 이상 진단 방법은 제조 공정의 센서(들)로부터 측정된 데이터를 수집하는 단계, 수집된 데이터 중 정상성을 나타내는 제1 데이터(S_0)와 실시간으로 수집되는 제2 데이터(S_n)를 구분하여 클래스를 부여하는 단계, 랜덤 포레스트를 이용하여 제1 데이터(S_0)와 제2 데이터(S_n)를 학습하고 의사결정트리의 예측 결과를 생성하는 단계, 예측 결과에 가중치를 적용하여 가중치가 적용된 분류 확률을 계산하는 단계, 제조 공정의 이상을 탐지하기 위한 모니터링 통계량을 생성하는 단계, 및 모니터링 통계량이 관리한계선을 벗어나는 경우 이상으로 탐지하는 단계를 포함한다.

대표도 - 도1



(52) CPC특허분류

G05B 23/0243 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호 F16SN26T4601

부처명 교육부

연구관리전문기관 한국연구재단

연구사업명 BK21플러스사업

연구과제명 제조 물류분야에서의 빅데이터 운용사업팀

기 여 율 1/2

주관기관 고려대학교 산학협력단

연구기간 2016.03.01 ~ 2017.02.28

이 발명을 지원한 국가연구개발사업

과제고유번호 1711040094

부처명 미래창조과학부

연구관리전문기관 한국연구재단

연구사업명 개인연구지원

연구과제명 빅 데이터 기반 스마트팩토리 구현을 위한 프리딕티브 에널리틱스 기법 개발

기 여 율 1/2

주관기관 고려대학교 산학협력단

연구기간 2016.06.01 ~ 2017.05.31

명세서

청구범위

청구항 1

개선된 RTC(Real-time contrasts) 기법을 이용한 지능형 이상 진단 장치에서 제조 공정의 이상을 탐지하고 원인을 진단하는 방법에 있어서,

제조 공정의 센서(들)로부터 측정된 데이터(x_i)를 수집하는 단계;

수집된 데이터 중 정상을 나타내는 제1 데이터(S_0)와 실시간으로 수집되는 제2 데이터(S_w)를 구분하여 클래스를 부여하는 단계;

랜덤 포레스트를 이용하여 상기 제1 데이터(S_0)와 상기 제2 데이터(S_w)를 학습하고 의사결정트리(T_j)의 예측 결과($\hat{y}(x_i, T_j)$)를 생성하는 단계;

상기 예측 결과($\hat{y}(x_i, T_j)$)에 가중치(w_j)를 적용하여 가중치가 적용된 분류 확률($\hat{p}_k(x_i)$)을 계산하는 단계;

상기 가중치가 적용된 분류 확률($\hat{p}_k(x_i)$)을 이용하여 제조 공정의 이상을 탐지하기 위한 모니터링 통계량($p(S_0, t)$)을 생성하는 단계; 및

상기 모니터링 통계량($p(S_0, t)$)이 관리한계선을 벗어나는 경우, 이상으로 탐지하는 단계;를 포함하는 개선된 RTC 기법을 이용한 지능형 이상 진단 방법.

청구항 2

제1항에 있어서,

상기 가중치가 적용된 분류 확률($\hat{p}_k(x_i)$)은 아래의 식을 이용하여 계산하는 개선된 RTC 기법을 이용한 지능형 이상 진단 방법.

$$\hat{p}_k(x_i) = \frac{\sum_{j \in OOB_i} w_j I[\hat{y}(x_i, T_j) = k]}{\sum_{j \in OOB_i} w_j}$$

(이때, OOB는 의사결정트리(T_j)를 학습할 때 사용되지 않은 데이터셋, $I(\cdot)$ 는 의사결정트리(T_j)의 예측결과($\hat{y}(x_i, T_j)$)가 실제 클래스 k 와 일치하면 1을, 일치하지 않으면 0을 반환하는 함수)

청구항 3

제1항에 있어서,

상기 모니터링 통계량($p(S_0, t)$)은 아래의 식을 이용하여 계산하는 개선된 RTC 기법을 이용한 지능형 이상 진단 방법.

$$p(S_0, t) = \frac{\sum_{x_i \in S_0} \hat{p}_k(x_i|t)}{N_0}$$

(이때, N_0 는 제1 데이터(S_0)의 사이즈)

청구항 4

제1항에 있어서,

신규의 데이터가 수집될 때마다, 클래스를 부여하는 단계, 의사결정트리(T_j)의 예측 결과($\hat{y}(x_i, T_j)$)를 생성하는 단계, 가중치가 적용된 분류 확률($\hat{p}_k(x_i)$)을 계산하는 단계, 및 모니터링 통계량($p(S_0, t)$)을 생성하는 단계를 반복하여 모니터링 통계량($p(S_0, t)$)을 갱신하는 개선된 RTC 기법을 이용한 지능형 이상 진단 방법.

청구항 5

제1항에 있어서,

상기 가중치(w_j)는 F-measure(F_j)를 이용한 제1 가중치(w_j^F), G-mean(G_j)을 이용한 제2 가중치(w_j^G), 또는 MCC(Matthews correlation coefficient, M_j)를 이용한 제3 가중치(w_j^M) 중 어느 하나인 개선된 RTC 기법을 이용한 지능형 이상 진단 방법.

청구항 6

제1항에 있어서,

지니 계수를 이용하여 이상의 원인이 되는 부분을 진단하는 단계를 더 포함하는 개선된 RTC 기법을 이용한 지능형 이상 진단 방법.

청구항 7

제조 공정의 센서(들)로부터 측정된 데이터(x_i)를 수집하고, 수집된 데이터 중 정상을 나타내는 제1 데이터(S_0)와 실시간으로 수집되는 제2 데이터(S_w)를 구분하여 클래스를 부여하고, 랜덤 포레스트를 이용하여 상기 제1 데이터(S_0)와 상기 제2 데이터(S_w)를 학습하고 의사결정트리(T_j)의 예측 결과($\hat{y}(x_i, T_j)$)를 생성하는 학습 모듈,

상기 예측 결과($\hat{y}(x_i, T_j)$)에 가중치(w_j)를 적용하여 가중치가 적용된 분류 확률($\hat{p}_k(x_i)$)을 계산하고, 상기 가중치가 적용된 분류 확률($\hat{p}_k(x_i)$)을 이용하여 제조 공정의 이상을 탐지하기 위한 모니터링 통계량($p(S_0, t)$)을 생성하여, 상기 모니터링 통계량($p(S_0, t)$)이 관리한계선을 벗어나는 경우, 이상으로 탐지하는 이상 탐지 모듈, 및

지니 계수를 이용하여 이상의 원인이 되는 부분을 진단하는 이상 진단 모듈을 포함하는 개선된 RTC 기법을 이용한 지능형 이상 진단 장치.

청구항 8

제7항에 있어서,

상기 가중치가 적용된 분류 확률($\hat{p}_k(x_i)$)은 아래의 식을 이용하여 계산하고,

$$\hat{p}_k(x_i) = \frac{\sum_{j \in OOB_i} w_j I[\hat{y}(x_i, T_j) = k]}{\sum_{j \in OOB_i} w_j}$$

(이때, OOB는 의사결정트리(T_j)를 학습할 때 사용되지 않은 데이터셋, $I(\cdot)$ 는 의사결정트리(T_j)의 예측결과($\hat{y}(x_i, T_j)$)가 실제 클래스 k 와 일치하면 1을, 일치하지 않으면 0을 반환하는 함수)

상기 모니터링 통계량($p(S_0, t)$)은 아래의 식을 이용하여 생성하는 ,

$$p(S_0, t) = \frac{\sum_{x_i \in S_0} \hat{p}_k(x_i|t)}{N_0}$$

(이때, N_0 는 제1 데이터(S_0)의 사이즈)

개선된 RTC 기법을 이용한 지능형 이상 진단 장치.

청구항 9

제7항에 있어서,

상기 가중치(w_j)는 F-measure(F_j)를 이용한 제1 가중치(w_j^F), G-mean(G_j)을 이용한 제2 가중치(w_j^G), 또는 MCC(Matthews correlation coefficient, M_j)를 이용한 제3 가중치(w_j^M) 중 어느 하나인 개선된 RTC 기법을 이용한 지능형 이상 진단 장치.

발명의 설명

기술 분야

[0001] 본 발명의 개념에 따른 실시 예는 개선된 RTC(Real-time contrasts) 기법을 이용한 지능형 이상 진단 방법 및 장치에 관한 것으로, 더욱 상세하게는, 제조공정에서 실시간으로 이상을 탐지하고, 이상의 원인을 실시간으로 파악하는 지능형 이상 진단 방법 및 장치에 관한 것이다.

배경 기술

[0002] 본 발명은 제품 제조 공정에서 제품의 이상을 탐지하고, 이상의 원인을 파악하는 지능형 이상 진단 기법에 관한 것이다.

[0003] 최근 센서 기술의 발달로 관리의 대상이 되는 시스템으로부터 실시간으로 데이터를 수집하는 것이 가능해졌다. 양질의 공정 데이터만 주어진다면 SPC(Statistical Process Control) 기법을 이용하여 비교적 쉽게 공정 상태를 모니터링할 수 있다. 그러나 실제 공정에서는 예상치 못한 공정 상태 변화, 비선형성, 비정규성 그리고 다중공선성 등으로 인해 기존의 SPC 기법을 이용한 공정 모니터링(process monitoring)은 비효율적이거나, 신뢰할 수 없는 결과를 야기하기도 한다.

[0004] RTC(Real-time contrasts) 기법은 공정 모니터링 문제를 실시간 분류(real-time classification) 문제로 전환한다. 이미 수집된 데이터 중 정상상을 나타내는 데이터(S_0 , reference data)에 클래스 0을 부여하고, 이후에 실시간으로 수집되는 데이터(S_1 , real time contrasts)에는 클래스 1을 부여한다. S_1 에는 크기가 w 인 이동식 윈도우(moving window)를 적용하여 S_1 이 가장 최근에 수집된 w 개의 데이터만을 포함하도록 한다. 데이터가 수집될 때마다 랜덤 포레스트(Random forests)를 이용하여 S_0 과 S_1 을 학습시킴으로써 분류 경계(classification boundary, decision boundary)를 갱신한다.

[0005] 기존의 RTC 기법은 랜덤 포레스트(Random forests)의 분류확률을 이용하는데, 이 분류확률은 랜덤 포레스트(Random forests)를 구성하는 의사결정나무 각각의 분류결과를 토대로 계산된다. 그 결과, 시스템이 더 큰 이상 징후를 갖고 있음에도 동일한 분류확률이 계산될 수 있고, 이로 인해 미세한 시스템의 이상을 탐지하지 못하는 경우가 발생한다. 또한, S_0 의 크기를 $Sw(t)$ 보다 크게 설정하는 경우 클래스 불균형(class imbalance) 문제가 발생하여 편향된 분류 경계(decision boundary)가 구축될 수 있는 문제점이 있다.

[0006] 따라서, 미세한 시스템의 이상을 효과적으로 탐지할 수 있고, 보다 정확한 분류 경계를 구축할 수 있는 지능형 이상 진단 기법이 요구된다.

선행기술문헌

비특허문헌

- [0007] (비특허문헌 0001) Deng, H., Runger, G., & Tuv, E. (2012). System monitoring with real-time contrasts. *Journal of Quality Technology*, 44(1), 9.
- (비특허문헌 0002) He, S., Jiang, W., & Deng, H. (2016). A distance-based control chart for monitoring multivariate processes using support vector machines. *Annals of Operations Research*, 1-17.
- (비특허문헌 0003) (특허문헌 3) Wei, Q., Huang, W., Jiang, W., & Zhao, W. (2016). Real-time process monitoring using kernel distances. *International Journal of Production Research*, 1-16.
- (비특허문헌 0004) (특허문헌 4) Kuncheva, L. I., & Rodriguez, J. J. (2014). A weighted voting framework for classifiers ensembles. *Knowledge and Information Systems*, 38(2), 259-275.

발명의 내용

해결하려는 과제

- [0008] 본 발명이 이루고자 하는 기술적인 과제는 기존의 RTC 기법을 개선하여 보다 효과적으로 시스템의 이상을 진단하는 방법을 제공하는 것이다.
- [0009] 본 발명이 이루고자 하는 기술적인 과제는 가중치 보팅(weighted voting)을 이용하여 트리 기반 분류기(tree-based classifier)의 monitoring statistics를 연속적으로 생성하고 이상 탐지 성능을 향상시키는 것이다.
- [0010] 또한 본 발명이 이루고자 하는 기술적인 과제는 가중치 보팅(weighted voting)을 이용하여 기존 RTC 기법의 class imbalance 문제를 해결하고 실시간으로 분류 경계를 업데이트 할 수 있는 지능형 이상 탐지 방법을 제공하는 것이다.

과제의 해결 수단

- [0011] 본 발명의 일 실시 예에 따른 개선된 RTC 기법을 이용한 지능형 이상 진단 방법은 제조 공정의 센서(들)로부터 측정된 데이터(x_i)를 수집하는 단계, 수집된 데이터 중 정상을 나타내는 제1 데이터(S_0)와 실시간으로 수집되는 제2 데이터(S_w)를 구분하여 클래스를 부여하는 단계, 랜덤 포레스트를 이용하여 상기 제1 데이터(S_0)와 상기 제2 데이터(S_w)를 학습하고 의사결정트리(T_j)의 예측 결과($\hat{y}(x_i, T_j)$)를 생성하는 단계, 상기 예측 결과($\hat{y}(x_i, T_j)$)에 가중치(w_j)를 적용하여 가중치가 적용된 분류 확률($\hat{p}_k(x_i)$)을 계산하는 단계, 상기 가중치가 적용된 분류 확률($\hat{p}_k(x_i)$)을 이용하여 제조 공정의 이상을 탐지하기 위한 모니터링 통계량($p(S_0, t)$)을 생성하는 단계, 및 상기 모니터링 통계량($p(S_0, t)$)이 관리한계선을 벗어나는 경우, 이상으로 탐지하는 단계를 포함한다.
- [0012] 본 발명의 일 실시 예에 따른 개선된 RTC 기법을 이용한 지능형 이상 진단 장치는 제조 공정의 센서(들)로부터 측정된 데이터(x_i)를 수집하고, 수집된 데이터 중 정상을 나타내는 제1 데이터(S_0)와 실시간으로 수집되는 제2 데이터(S_w)를 구분하여 클래스를 부여하고, 랜덤 포레스트를 이용하여 상기 제1 데이터(S_0)와 상기 제2 데이터(S_w)를 학습하고 의사결정트리(T_j)의 예측 결과($\hat{y}(x_i, T_j)$)를 생성하는 학습 모듈, 상기 예측 결과($\hat{y}(x_i, T_j)$)에 가중치(w_j)를 적용하여 가중치가 적용된 분류 확률($\hat{p}_k(x_i)$)을 계산하고, 상기 가중치가 적용된 분류 확률($\hat{p}_k(x_i)$)을 이용하여 제조 공정의 이상을 탐지하기 위한 모니터링 통계량($p(S_0, t)$)을 생성하여, 상기 모니터링 통계량($p(S_0, t)$)이 관리한계선을 벗어나는 경우, 이상으로 탐지하는 이상 탐지 모듈, 및 지니 계수를 이용하여 이상의 원인이 되는 부분을 진단하는 이상 진단 모듈을 포함한다.

발명의 효과

- [0013] 본 발명의 실시 예에 따른 개선된 RTC 기법을 이용한 지능형 이상 진단 방법 및 장치는 랜덤 포레스트(random forests)를 구성하는 의사결정나무 각각의 분류결과에 가중치 보팅(weighted voting) 방식을 사용함으로써 최종 분류확률 값을 연속(continuous)하게 만들어 미세한 시스템의 이상도 효과적으로 탐지할 수 있는 효과가 있다.
- [0014] 본 발명의 실시 예에 따른 개선된 RTC 기법을 이용한 지능형 이상 진단 방법 및 장치는 랜덤 포레스트(Random forests)를 구성하는 의사결정나무의 분류정확도를 평가하는 척도로 F-measure, G-mean, MCC(Matthews Correlation Coefficient)를 사용하여 가중치(weight)를 부여함으로써, S0와 S1의 크기가 달라 발생하는 클래스 불균형 문제를 해소하고, 이에 따라 기존 RTC 기법에서 생성된 편향(bias)된 분류 경계를 개선하여 보다 정확한 분류 결과를 획득할 수 있는 효과가 있다.
- [0015] 또한, 본 발명의 실시 예에 따른 개선된 RTC 기법을 이용한 지능형 이상 진단 방법 및 장치는 랜덤 포레스트(Random forests)로부터 계산되는 지니 중요도(Gini importance)를 이용하여 어느 변수에서 이상이 발생하였는지를 파악할 수 있는 효과가 있다.

도면의 간단한 설명

- [0016] 본 발명의 상세한 설명에서 인용되는 도면을 보다 충분히 이해하기 위하여 각 도면의 상세한 설명이 제공된다.
- 도 1은 본 발명의 일 실시 예에 따른 개선된 RTC 기법을 이용한 지능형 이상 진단 장치의 기능 블록도이다.
- 도 2는 본 발명의 일 실시 예에 따른 개선된 RTC 기법을 이용한 지능형 이상 진단 방법을 설명하기 위한 흐름도이다.
- 도 3 및 도 4는 본 발명의 일 실시예에 따른 개선된 RTC(Real-time contrasts) 기법을 이용한 모니터링 통계량 및 기존의 기법을 이용한 모니터링 통계량을 비교하여 도시한 예시적인 그래프이다.
- 도 5 내지 도 8은 정규 분포를 따르는 데이터에 대하여 네 가지 차원(10, 20, 50, 100)에 따른 이상 탐지 결과를 도시한 표이다.
- 도 9 내지 도 10은 비정규 분포(t 분포, 감마 분포)를 따르는 데이터에 대하여 시프트 사이즈를 달리하여 이상 탐지를 수행한 결과를 도시한 표이다.
- 도 11 및 도 12는 본 발명의 일 실시예에 따른 개선된 RTC(Real-time contrasts) 기법을 이용한 지능형 이상 진단 방법으로 이상을 탐지하고 이상의 원인을 진단한 예시적인 그래프이다.

발명을 실시하기 위한 구체적인 내용

- [0017] 본 명세서에 개시되어 있는 본 발명의 개념에 따른 실시 예들에 대해서 특정한 구조적 또는 기능적 설명은 단지 본 발명의 개념에 따른 실시 예들을 설명하기 위한 목적으로 예시된 것으로서, 본 발명의 개념에 따른 실시 예들은 다양한 형태들로 실시될 수 있으며 본 명세서에 설명된 실시 예들에 한정되지 않는다.
- [0018] 본 발명의 개념에 따른 실시 예들은 다양한 변경들을 가할 수 있고 여러 가지 형태들을 가질 수 있으므로 실시 예들을 도면에 예시하고 본 명세서에서 상세하게 설명하고자 한다. 그러나 이는 본 발명의 개념에 따른 실시 예들을 특정한 개시 형태들에 대해 한정하려는 것이 아니며, 본 발명의 사상 및 기술 범위에 포함되는 모든 변경, 균등물, 또는 대체물을 포함한다.
- [0019] 본 명세서에서 사용한 용어는 단지 특정한 실시 예를 설명하기 위해 사용된 것으로서, 본 발명을 한정하려는 의도가 아니다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 명세서에서, "포함하다" 또는 "가지다" 등의 용어는 본 명세서에 기재된 특징, 숫자, 단계, 동작, 구성 요소, 부분품 또는 이들을 조합한 것이 존재함을 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성 요소, 부분품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.
- [0020] 다르게 정의되지 않는 한, 기술적이거나 과학적인 용어를 포함해서 여기서 사용되는 모든 용어들은 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미를 가진다. 일반적으로 사용되는 사전에 정의되어 있는 것과 같은 용어들은 관련 기술의 문맥상 가지는 의미와 일치하는 의미를 갖는 것으로 해석되어야 하며, 본 명세서에서 명백하게 정의하지 않는 한, 이상적이거나 과도하게 형식적인 의

미로 해석되지 않는다.

- [0021] 이하, 본 명세서에 첨부된 도면들을 참조하여 본 발명의 실시 예들을 상세히 설명한다.
- [0022] 먼저, 도 1을 참조하여, 본 발명의 일 실시예에 따른 개선된 RTC(Real-time contrasts) 기법을 이용한 지능형 이상 진단 장치에 대해 상술한다.
- [0023] 도 1은 본 발명의 일 실시 예에 따른 개선된 RTC(Real-time contrasts) 기법을 이용한 지능형 이상 진단 장치의 기능 블록도이다. 도 1을 참조하면, 지능형 이상 진단 장치(10)는 학습 모듈(100), 이상 탐지 모듈(200), 이상의 원인을 파악하는 원인 진단 모듈(300), 및 데이터베이스(800)을 포함한다.
- [0024] 본 명세서에서 사용되는 '-부' 또는 '모듈'이라 함은 본 발명의 기술적 사상을 수행하기 위한 하드웨어 및 상기 하드웨어를 구동하기 위한 소프트웨어의 기능적, 구조적 결합을 의미할 수 있다. 예컨대, 상기 '-부' 또는 '-모듈'은 소정의 코드와 상기 소정의 코드가 수행되기 위한 하드웨어 리소스의 논리적인 단위를 의미할 수 있으며, 반드시 물리적으로 연결된 코드를 의미하거나 한 종류의 하드웨어를 의미하는 것은 아니다.
- [0025] 지능형 이상 진단 장치(10)의 학습 모듈(100)은 클래스 구분부(110) 및 개별 분류기 학습부(130)를 포함한다.
- [0026] 학습 모듈(100)의 클래스 구분부(110)는 공정 데이터(x_i)를 레퍼런스(reference) 데이터 S_0 와 RTC(real-time contrasts) 데이터 $S_w(t)$, 두 개의 클래스로 구분한다. 구체적으로, 특정 제품을 제조하는 공정에서 센서(들)로부터 측정되고 수집된 데이터(x_i) 중 정상상을 나타내는 데이터(S_0)에 클래스 0을 부여하고, 이후에 실시간으로 수집되는 데이터($S_w(t)$)에는 클래스 1을 부여한다. 이때, $S_w(t)$ 에는 크기가 N_w 인 이동식 윈도우(moving window)를 적용하여, $S_w(t)$ 가 가장 최근에 수집된 N_w 개의 데이터만을 포함하도록 한다.
- [0027] 학습 모듈(100)의 개별 분류기 학습부(130)는 클래스가 부여된 레퍼런스 데이터 및 RTC 데이터를 이용하여 개별 의사결정트리(T_j)로 데이터를 학습하여 예측결과 $\hat{y}(x_i, T_j)$ 를 생성한다. 이때, j 는 개별 분류기의 인덱스를 의미하며, T_j 는 j 번째 분류기(의사결정트리)를 의미한다.
- [0028] 개별 분류기 학습부(130)는 신규 데이터를 수집할 때마다, 즉 무빙 윈도우에 들어오는 데이터가 업데이트될 때마다, 분류기(의사결정트리)를 새로 생성한다.
- [0029] 지능형 이상 진단 장치(10)의 이상 탐지 모듈(200)은 가중치 결정부(210), 가중치 적용 분류 확률 생성부(230), 모니터링 통계량 생성부(250), 이상탐지기준결정부(270) 및 이상탐지부(270)를 포함한다.
- [0030] 이상 탐지 모듈(200)은 가중치 결정부(210)는 가중치가 적용된 분류 확률을 계산하기 위하여 개별 의사결정트리의 분류 결과 $\hat{y}(x_i, T_j)$ 에 적용할 가중치(w_j)를 결정한다.
- [0031] 본 발명에서는 소수 클래스($S_w(t)$)에 집중한 평균값을 이용한 가중치를 적용한다. 구체적으로, 가중치가 적용된 분류 확률의 계산에 적용되는 가중치는 F-measure(F_j)를 이용한 제1 가중치(w_j^F), G-mean(G_j)을 이용한 제2 가중치(w_j^G), 또는 MCC(Matthews correlation coefficient, M_j)를 이용한 제3 가중치(w_j^M) 중 어느 하나일 수 있다. 가중치의 상세한 설명은 지능형 이상 진단 방법에서 가중치에 대한 설명과 동일하므로 생략한다.
- [0032] 이상 탐지 모듈(200)의 가중치 적용 분류 확률 생성부(230)는 가중치(w_j)가 적용된 랜덤 포레스트(random forests) 분류 확률을 계산한다(S210). 가중치(w_j)가 적용된 랜덤 포레스트(random forests) 분류 확률 $\hat{p}_k(x_i)$ 는 아래의 식을 이용하여 계산한다.

$$\hat{p}_k(x_i) = \frac{\sum_{j \in OOB_i} w_j I[\hat{y}(x_i, T_j) = k]}{\sum_{j \in OOB_i} w_j}$$

- [0033]
- [0034] 랜덤 포레스트(random forests)를 구성하는 개별 의사결정트리(decision tree)들의 분류 결과에 가중치(weight)를 적용한 것으로 과적합(Overfitting)을 방지하기 위하여 OOB(Out of bagging)를 이용하여 분류확률

을 계산한다. OOB는 개별 의사결정트리(decision tree)를 학습할 때 사용되지 않은 데이터셋(Dataset)을 의미한다. 예를 들어, 랜덤 포레스트로 데이터 학습하기 위하여 부트스트랩 기법을 이용하여 데이터를 랜덤으로 추출하는 경우, OOB는 의사결정트리를 생성할 때 추출하지 않은 데이터를 의미한다. $I(\cdot)$ 는 개별 의사결정트리(decision tree)의 예측결과 $\hat{y}(x_i, T_j)$ 가 실제클래스 k와 일치하면 1을, 일치하지 않으면 0을 반환하는 함수이다.

[0035] 이상 탐지 모듈(200)의 모니터링 통계량 생성부(250)는 가중치가 적용된 분류 확률($\hat{p}_k(x_i)$)을 이용하여 이상 탐지를 위한 컨트롤 차트의 모니터링 통계량($p(S_0, t)$)을 생성한다. 모니터링 통계량($p(S_0, t)$)은 정확하고 안정적인 결과를 도출하기 위하여 S_0 를 사용하여, 아래의 식과 같이 생성할 수 있다.

$$p(S_0, t) = \frac{\sum_{x_i \in S_0} \hat{p}_k(x_i | t)}{N_0}$$

[0037] 이때, N_0 는 제1 데이터(S_0)의 사이즈를 의미한다.

[0038] 이상 탐지 모듈(200)의 이상탐지기준 결정부(270)는 사용자 또는 관리자 등에 의하여 관리한계선(Control Limit)을 결정한다. 이상탐지기준은 탐지하고자 하는 이상에 따라, 이상 탐지 대상 제품에 따라, 또는 이상 탐지 대상 제품의 공정 환경 등에 따라, 이상탐지기준이 되는 관리한계선을 결정할 수 있다.

[0039] 예를 들어, 또한, 모니터링 통계량($p(S_0, t)$)의 관리한계선은 사용자가 정한 ARL(average run length)을 이용하여 계산할 수 있다. ARL은 관리도의 성능을 평가하는 척도 중 하나로서, 관리도가 이상상태를 감지하기 전까지의 기대 점의 개수이다.

[0040] 이상 탐지 모듈(200)의 이상 탐지부(290)은 모니터링 통계량이 관리한계선을 벗어난 경우 제조 공정에 이상이 발생한 것으로 판단한다. 이상이 탐지된 경우 이상 탐지부(290)은 이상 탐지 정보를 데이터베이스(800)에 저장하고 출력할 수 있다.

[0041] 지능형 이상 진단 장치(10)의 원인 진단 모듈(300)은 불순도 계산부(310) 및 이상 원인 진단부(330)를 포함한다.

[0042] 원인 진단 모듈(300)의 불순도 계산부(310)는 이상 원인을 진단하기 위하여 불순도(impurity, $I(v)$)를 계산한다. 예를 들어, 불순도는 지니 중요도(Gini Importance, 지니 계수)일 수 있다.

[0043] 이상 원인 진단부(330)는 불순도($I(v)$, 예를 들어, 지니 계수)를 이용하여 이상의 원인이 되는 부분을 진단한다. 구체적으로, 아래의 식과 같이 각 변수(m)가 이상을 탐지하는 데 기여하는 정도($Imp(X^m)$)를 계산하여, m번째 변수의 $Imp(X^m)$ 값이 크면, m번째 변수에 이상의 원인이 있는 것으로 판단한다.

$$Imp(X^m) = \frac{1}{ntree} \sum_{j=1}^{ntree} \sum_{v \in T_j: X^m} \Delta I(s_v, v)$$

$$, \Delta I(s_v, v) = I(v) - (w_L I(v_L) + w_R I(v_R))$$

[0045] 데이터베이스(800)는 수집된 데이터, 개별 의사결정트리, 가중치, 분류 확률, 모니터링 통계량, 이상 탐지 기준, 이상 탐지 결과, 또는 이상 원인 진단 결과 등을 저장할 수 있다. 본 명세서에서 데이터베이스라 함은, 각각의 데이터베이스에 대응되는 정보를 저장하는 소프트웨어 및 하드웨어의 기능적 구조적 결합을 의미할 수도 있다.

[0046] 제어모듈(미도시)은 지능형 이상 진단 장치(10)의 전반적인 동작을 제어한다. 즉, 학습 모듈(100), 이상 탐지 모듈(200), 원인 진단 모듈(300), 및 데이터베이스(800)의 동작을 제어할 수 있다.

[0047] 본 실시예와 달리, 상기 지능형 이상 진단 장치(10)가 아닌 별도의 장치에 학습 모듈(100), 이상 탐지 모듈(200), 원인 진단 모듈(300), 또는 데이터베이스(800)) 중 어느 하나 이상을 구비하는 것도 가능하다. 이때, 각각의 장치별로 제어모듈을 각각 구비할 수 있다.

[0048] 이하, 도 2 내지 도 12을 참조하여, 본 발명의 일 실시예에 따른 개선된 RTC(Real-time contrasts) 기법을 이용

한 지능형 이상 진단 장치를 이용한 개선된 RTC(Real-time contrasts) 기법을 이용한 지능형 이상 진단 방법에 대하여 자세히 살펴보도록 한다.

[0049] 도 2는 도 1에 도시한 개선된 RTC(Real-time contrasts) 기법을 이용한 지능형 이상 진단 장치를 이용한 개선된 RTC(Real-time contrasts) 기법을 이용한 지능형 이상 진단 방법을 설명하기 위한 흐름도이다.

[0050] 도 2를 참조하면, 개선된 RTC(Real-time contrasts) 기법을 이용한 지능형 이상 진단 장치를 이용한 개선된 RTC(Real-time contrasts) 기법을 이용한 지능형 이상 진단 방법은 센서에서 수집되는 데이터로부터 정상 데이터와 실시간으로 수집되는 데이터를 추출하여 각각 클래스 0, 1을 부여하고 개별 의사결정나무로 데이터를 학습하는 학습 단계(S100), 학습된 개별 결정나무에 가중치 보팅을 적용하여 최종 분류 확률을 획득하고 분류 확률을 이용하여 시스템의 이상을 탐지하는 이상 탐지 단계(S200), 및 학습된 개별 의사결정나무로부터 각 변수의 지니 중요도를 계산하여 그 값을 비교함으로써 이상 원인을 파악하는 원인 진단 단계(S300)을 포함한다.

[0051] 먼저, 지능형 이상 진단 장치(10)의 학습 모듈(100)은 특정 제품을 제조하는 공정에서, 센서(들)로부터 측정된 데이터(x_i)를 수집한다(S110).

[0052] 다음, 수집된 데이터 중 정상을 나타내는 데이터(S_0)에 클래스 0을 부여하고, 이후에 실시간으로 수집되는 데이터($S_w(t)$)에는 클래스 1을 부여한다(S130). 즉, 레퍼런스(reference) 데이터 S_0 와 RTC(real-time contrasts) 데이터 $S_w(t)$ 는 두 개의 클래스로 구분한다.

[0053] $S_w(t)$ 에는 크기가 N_w 인 이동식 윈도우(moving window)를 적용하여, 아래의 [수학식 1]과 같이 $S_w(t)$ 가 가장 최근에 수집된 N_w 개의 데이터만을 포함하도록 한다.

수학식 1

[0054]
$$S_w(t) = \{x_t - N_w + 1, \dots, x_t - 1, x_t\}$$

[0055] 다음, 클래스가 부여된 레퍼런스 데이터 및 RTC 데이터를 이용하여 개별 의사결정트리로 데이터를 학습한다(S150). 즉, 의사결정트리(decision tree)의 예측결과 $\hat{y}(x_i, T_j)$ 를 생성한다. 이때, j 는 개별 분류기의 인덱스를 의미하며, T_j 는 j 번째 분류기(의사결정트리)를 의미한다.

[0056] 이때, 신규 데이터를 수집할 때마다, 즉 무빙 윈도우에 들어오는 데이터가 업데이트될 때마다, 분류기(의사결정트리)를 새로 생성한다.

[0057] 다음, 지능형 이상 진단 장치(10)은 가중치(w_j)가 적용된 랜덤 포레스트(random forests) 분류 확률을 계산한다(S210). 가중치(w_j)가 적용된 랜덤 포레스트(random forests) 분류 확률 $\hat{p}_k(x_i)$ 는 아래의 [수학식 2]를 이용하여 계산한다.

수학식 2

[0058]
$$\hat{p}_k(x_i) = \frac{\sum_{j \in OOB_i} w_j I[\hat{y}(x_i, T_j) = k]}{\sum_{j \in OOB_i} w_j}$$

[0059] 랜덤 포레스트(random forests)를 구성하는 개별 의사결정트리(decision tree)들의 분류 결과에 가중치(weight)를 적용한 것으로 과적합(Overfitting)을 방지하기 위하여 OOB(Out of bagging)를 이용하여 분류확률을 계산한다. OOB는 개별 의사결정트리(decision tree)를 학습할 때 사용되지 않은 데이터셋(Dataset)을 의미한다. 예를 들어, 랜덤 포레스트로 데이터 학습하기 위하여 부트스트랩 기법을 이용하여 데이터를 랜덤으로 추출하는 경우, OOB는 의사결정트리를 생성할 때 추출하지 않은 데이터를 의미한다.

[0060] $I(\cdot)$ 는 개별 의사결정트리(decision tree)의 예측결과 $\hat{y}(x_i, T_j)$ 가 실제클래스 k와 일치하면 1을, 일치하지 않으면 0을 반환하는 함수이다.

[0061] 즉, [수학식 2]를 이용하여 개별 의사결정트리(decision tree)의 분류결과에 가중치를 부여하여 랜덤 포레스트(random forests)의 옳게 분류할 확률(각 분류기의 제대로 분류할 확률)을 계산한다.

[0062] 본 발명에서는 소수 클래스(Sw(t))에 집중한 평균값을 이용한 가중치를 이용한다. 구체적으로, 가중치가 적용된 분류 확률을 계산에 있어서, F-measure(F_j)를 이용한 제1 가중치(w_j^F), G-mean(G_j)을 이용한 제2 가중치(w_j^G), 또는 MCC(Matthews correlation coefficient, M_j)를 이용한 제3 가중치(w_j^M) 중 어느 하나의 가중치(w_j)를 적용한다.

[0063] 제1 가중치(w_j^F)는 아래의 [수학식 3]을 이용하여 계산한다.

수학식 3

$$w_j^F = \log \frac{1+F_j}{1-F_j}$$

[0065] F-measure(F_j)는 아래의 [수학식 4]를 이용하여 계산한다.

수학식 4

$$\text{F-measure} = \frac{2 \times \text{recall} \times \text{precision}}{\text{recall} + \text{precision}} = \frac{2TP}{2TP + FP + FN}$$

[0067] 이때, TP(True positive)는 실제 이상인 것을 이상으로 분류한 것, TN(True negative)는 실제 정상인 것을 정상으로 분류한 것, FP(False positive)는 실제로는 정상인데 이상으로 오분류한 것, FN(False negative)는 실제로는 이상인데 정상으로 오분류한 것을 의미한다.

[0068] 제2 가중치(w_j^G)는 아래의 [수학식 5]를 이용하여 계산한다.

수학식 5

$$w_j^F = \log \frac{1+F_j}{1-F_j}$$

[0070] G-mean(G_j)는 아래의 [수학식 6]을 이용하여 계산한다.

수학식 6

$$\begin{aligned} \text{G-mean} &= \sqrt{\text{sensitivity} \times \text{specificity}} \\ &= \sqrt{\frac{TP \times TN}{(TP + FN)(TN + FP)}} \end{aligned}$$

[0072] 제3 가중치(w_j^M)는 아래의 [수학식 7]을 이용하여 계산한다.

수학식 7

$$w_j^M = \log \frac{1+M_j}{1-M_j}$$

[0073]

[0074] MCC(Matthews correlation coefficient, M_j)는 아래의 [수학식 8]을 이용하여 계산한다.

수학식 8

$$\begin{aligned} \text{MCC} &= \sqrt{\frac{(\text{sensitivity} + \text{specificity} - 1)(\text{precision} + \text{NPV} - 1)}{TP \times TN - FP \times FN}} \\ &= \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \end{aligned}$$

[0075]

[0076] 클래스 불균형(class imbalance)의 문제점이 존재하는 RTC 기법(RTC 데이터($S_w(t)$)가 레퍼런스 데이터(S_0)보다 적음)을 적용함에 있어서, 정상 또는 이상을 제대로 분류할 종래의 정확도(accuracy = $(TP+TN)/(TP+TN+FP+FN)$)의 경우 다수 클래스(S_0)에 편중된 수치를 보이는 문제점이 존재한다. 따라서, 소수 클래스에 집중한 F-measure(F_j), G-mean(G_j), 및 MCC(Matthews correlation coefficient, M_j)를 사용하는 경우 소수 클래스의 예측 결과에 보다 집중할 수 있는 효과가 있으며, 다수 클래스에 편중된 수치를 보여 분류기가 제대로 구축되었는지 평가하기 어려운 문제점을 해결할 수 있다.

[0077] F-measure는 G-mean이나 MCC보다 소수클래스에 집중한 수치를 계산하고, G-mean과 MCC는 소수클래스와 다수클래스의 예측결과를 동등하게 반영한 수치를 계산한다.

[0078] 본 발명에 따른 가중치(w_j)를 각 분류기에 적용함에 있어서, 각 분류기가 서로 독립이라는 가정하에 최적의 분류 성능을 얻을 수 있는 효과가 있다.

[0079] 다음, 가중치(w_j)가 적용된 랜덤 포레스트(random forests) 분류 확률($\hat{p}_k(x_i)$)을 이용하여 모니터링 통계량(monitored statistics, $p(S_0, t)$)을 생성한다(S230). 모니터링 통계량은 분류 확률($\hat{p}_k(x_i)$)의 평균일 수 있다. 구체적으로, 제1 가중치, 제2 가중치, 또는 제3 가중치 중 어느 하나의 가중치가 적용된 분류 확률 값을 이용하여 제조 공정 등의 이상을 탐지하기 위한 컨트롤 차트(control chart)의 모니터링 통계량을 아래의 [수학식 9]를 이용하여 계산한다.

수학식 9

$$p(S_0, t) = \frac{\sum_{x_i \in S_0} \hat{p}_k(x_i|t)}{N_0}$$

[0080]

[0081] 이때, 정확하고 안정적인 결과를 도출하기 위하여 S_0 를 사용한다.

[0082] 지능형 이상 진단 장치(10)는 새로운 데이터가 수집될 때마다 랜덤 포레스트를 이용하여 레퍼런스 데이터 및 RTC 데이터를 학습하여 가중치가 적용된 분류 확률 및 이를 이용한 모니터링 통계량을 갱신한다.

[0083] 다음, 관리자 또는 사용자 등에 의하여 미리 설정된 관리한계선(Control Limit:CL)을 이용하여 제조 공정 등의 이상을 탐지한다(S250). 즉, 모니터링 통계량이 관리한계선을 벗어나는 경우 이상으로 판단한다.

[0084] 관리한계선은 탐지하고자 하는 이상에 따라, 이상 탐지 대상 제품에 따라, 또는 이상 탐지 대상 제품의 공정 환경 등에 따라, 관리자 또는 사용자 등이 결정할 수 있다.

- [0085] 또한, 모니터링 통계량($p(S_0, t)$)의 관리한계선은 사용자가 정한 ARL(average run length)을 이용하여 계산할 수 있다. ARL은 관리도의 성능을 평가하는 척도 중 하나로, 관리도가 이상상태를 감지하기 전까지의 기대 점의 개수이다.
- [0086] ARL은 제1 ARL 또는 제2 ARL 중 어느 하나일 수 있다. 제1 ARL(In control-ARL, ARL_0)은 공정이 현재 관리 상태일 때 처음으로 이상 상태를 감지할 때까지의 기대점의 개수를 의미하고, 제2 ARL(out-of-control-ARL, ARL_1)은 공정이 현재 이상 상태일 때 처음으로 이상 상태를 감지할 때까지의 기대 점의 개수이다.
- [0087] 예를 들어, 제1 ARL(ARL_0)이 200에 근사하도록 시뮬레이션 실험을 통해 관리한계선을 결정할 수 있다.
- [0088] 이상이 탐지된 경우, 이상 탐지 모듈(300)은 이상 탐지 정보를 출력하고, 이상 탐지 결과를 데이터베이스(800)에 저장할 수 있다.
- [0089] 도 3 및 도 4는 비정규 분포의 데이터에 대하여 본 발명의 일 실시예에 따른 개선된 RTC(Real-time contrasts) 기법을 이용한 모니터링 통계량 및 기존의 기법을 이용한 모니터링 통계량을 비교하여 시뮬레이션한 예시적인 그래프이다. 무빙 윈도우 사이즈 및 참조 데이터 사이즈를 각각 10 및 1000으로 설정하고, 시프트 사이즈(shift size)를 1.5(도 3) 및 0.5(도 4)로 설정하였을 때의 모니터링 통계량 및 이를 이용하여 이상을 탐지한 도면이다. 다이아몬드 점이 최초로 관리한계선을 넘어섰다는 것을 의미하며, 본 발명의 제1 가중치를 적용한 모니터링 통계량($FWRF-p(S_0, t)$), 제2 가중치를 적용한 모니터링 통계량($GWRE-p(S_0, t)$), 또는 제3 가중치를 적용한 모니터링 통계량($MWRF-p(S_0, t)$)의 경우 가중치를 적용하지 않은 종래의 기법에 의한 모니터링 통계량($RF-p(S_0, t)$)에 비하여 보다 민감하게 공정 상황의 변화를 반영한다.
- [0090] 도 5 내지 도 8은 정규 분포를 따르는 데이터에 대하여 네 가지 차원(10, 20, 50, 100)에 따른 이상 탐지 결과를 도시한 표이다. 시나리오 1 복수의 변수가 하나의 표준 편차에 의하여 시프트(shift)된 시나리오이고, 시나리오 2는 다른 크기로 하나의 변수가 시프트된 시나리오이다. 수치가 낮은 것이 이상을 빠르게 탐지했다는 것을 의미한다.
- [0091] 도 5 내지 도 8을 참조하면, 데이터가 정규분포를 따를 때, 제1 가중치를 적용한 모니터링 통계량($FWRE-p(S_0, t)$), 또는 제2 가중치를 적용한 모니터링 통계량($GWRF-p(S_0, t)$)이 비교적 좋은 성능을 보이며, 데이터의 차원이 낮을 때는 제1 가중치를 적용한 모니터링 통계량($FWRF-p(S_0, t)$)이 데이터의 차원이 높을 때는 제3 가중치를 적용한 모니터링 통계량($MWRF-p(S_0, t)$)이 좋은 성능을 보인다.
- [0092] 도 9 내지 도 10은 비정규 분포(t 분포, 감마 분포)를 따르는 데이터에 대하여 시프트 사이즈를 달리하여 이상 탐지를 수행한 결과를 도시한 표이다. 수치가 낮은 것이 이상을 빠르게 탐지했다는 것을 의미한다.
- [0093] 도 9 내지 도 10을 참조하면, 데이터가 정규분포를 따르지 않을 때(t분포, 감마분포)는 제2 가중치를 적용한 모니터링 통계량($GWRF-p(S_0, t)$)이 좋은 성능을 보인다.
- [0094] 본 발명에 따른 이상 탐지 방법은 데이터가 정규분포를 따르지 않을 때, 데이터의 차원이 높을 때, 시프트 사이즈(shift size)가 작을 때, 특히 뛰어난 성능을 보인다. 즉, 본 발명에 따른 가중치를 적용한 모니터링 통계량은 복잡한 공정에서 미세한 시프트(shift, 이상)가 발생한 경우 효과적으로 이상을 탐지할 수 있다.
- [0095] 다음, 이상 원인을 진단하기 위하여 지니 계수를 결정한다(S310). 지니 계수($I(v)$)는 아래의 [수학식 10]을 이용하여 계산한다.

수학식 10

$$I(v) = \sum_{k=1}^C r_k(1 - r_k)$$

[0097] r_k 는 각 노드에서의 클래스 k의 비율을 나타낸다.

[0098] 다음, 지니 계수를 이용하여 이상의 원인이 되는 부분을 진단한다(S330). 구체적으로, 아래의 [수학식 11]을 이용하여, 각 변수(m)가 이상을 탐지하는 데 기여하는 정도($Imp(X^m)$)를 계산하여, m번째 변수의 $Imp(X^m)$ 값이 크면, m번째 변수에 이상의 원인이 있는 것으로 판단한다.

수학식 11

$$Imp(X^m) = \frac{1}{ntree} \sum_{j=1}^{ntree} \sum_{v \in T_j: X^m} \Delta I(s_v, v)$$

$$, \Delta I(s_v, v) = I(v) - (w_L I(v_L) + w_R I(v_R))$$

[0099]

[0100] 각 의사결정트리(decision tree)를 구성하는 모든 노드에 대하여 $\Delta I(s_v, v)$ 의 평균을 계산하여 $Imp(X^m)$ 값을 획득한다. $\Delta I(s_v, v)$ 는 의사결정트리의 각 노드에서 계산되는데, 노드에서 분기가 일어나기 전(부모 노드)의 $I(v)$ 와 분기가 일어난 후(자식 노드)의 좌우 노드에서의 $(w_L I(v_L) + w_R I(v_R))$ 과의 차이를 계산하여 해당 노드에서의 분기가 얼마나 불순도(impurity)를 줄여 잘 분류했는지를 계산한다. 이때, ntree는 의사결정트리의 개수이다.

[0101] 본 발명에서는 불순도($I(v)$)에 지니(gini) 계수를 사용하였으나, 이에 한정되지 아니하며 불순도(impurity)를 계산하는 다양한 계수(기법)들을 사용할 수 있다.

[0102] 도 11 및 도 12는 본 발명의 일 실시예에 따른 개선된 RTC(Real-time contrasts) 기법을 이용한 지능형 이상 진단 방법으로 이상을 탐지하고 이상의 원인을 진단한 예시적인 그래프이다. 도 11은 10개의 변수 중 하나의 변수에 이상이 있을 때, 이상 원인이 되는 변수(X1)를 탐지한 그래프이고, 도 12는 10개의 변수 중 세 개의 변수에 이상이 있을 때, 이상 원인이 되는 변수들(X1, X2, X3)를 탐지한 그래프로, 본 발명에 따른 지능형 이상 진단 방법을 통하여 이상의 원인을 정확하게 진단할 수 있다.

[0103] 본 발명은 도면에 도시된 실시 예를 참고로 설명되었으나 이는 예시적인 것에 불과하며, 본 기술 분야의 통상의 지식을 가진 자라면 이로부터 다양한 변형 및 균등한 타 실시 예가 가능하다는 점을 이해할 것이다. 따라서, 본 발명의 진정한 기술적 보호 범위는 첨부된 등록청구범위의 기술적 사상에 의해 정해져야 할 것이다.

부호의 설명

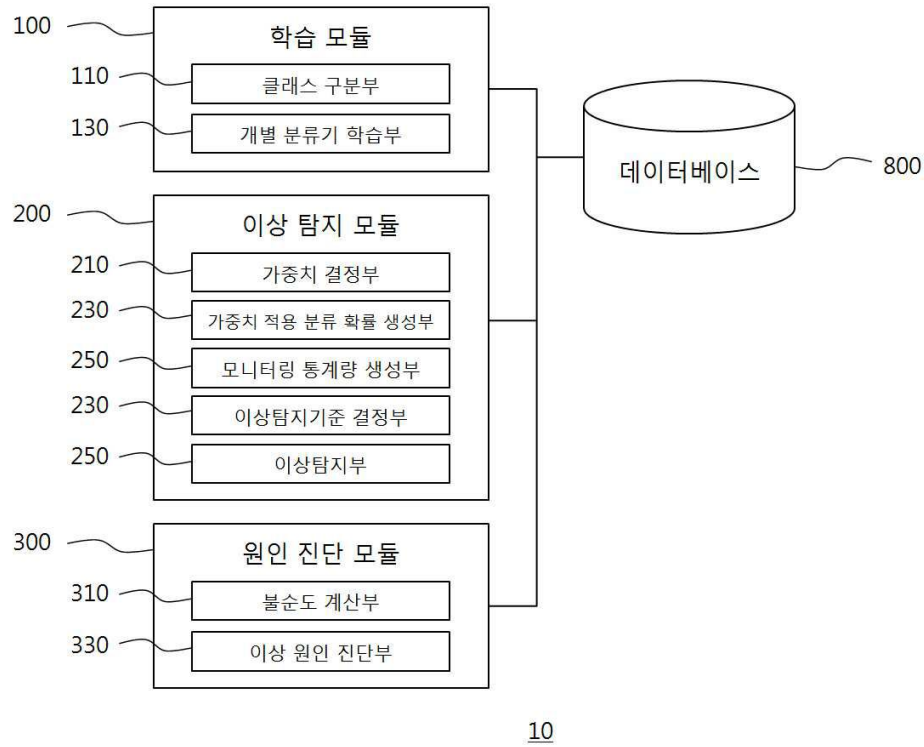
[0104] 10 : 개선된 RTC(기법을 이용한 지능형 이상 진단 장치

100 : 학습 모듈 200 : 이상 탐지 모듈

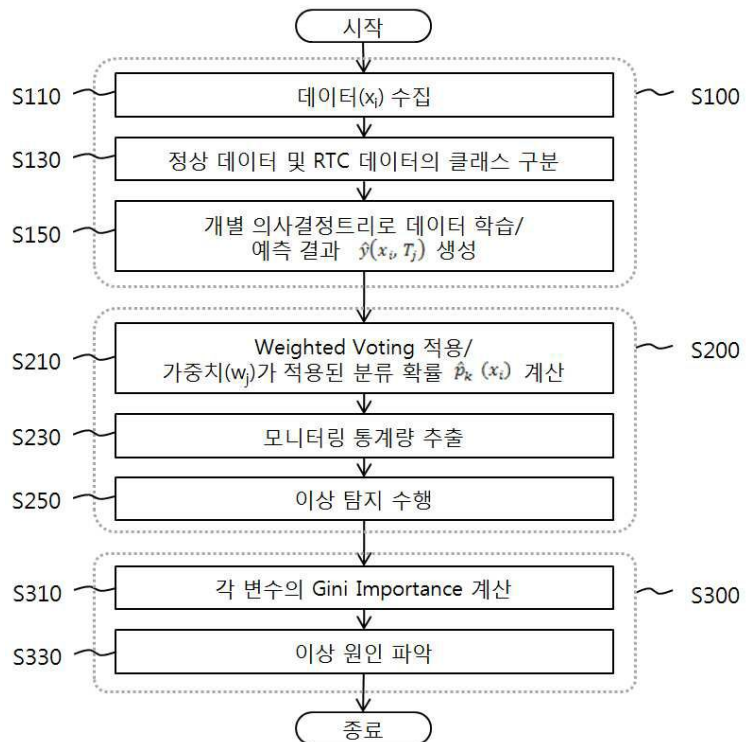
300 : 이상 진단 모듈 800 : 데이터베이스

도면

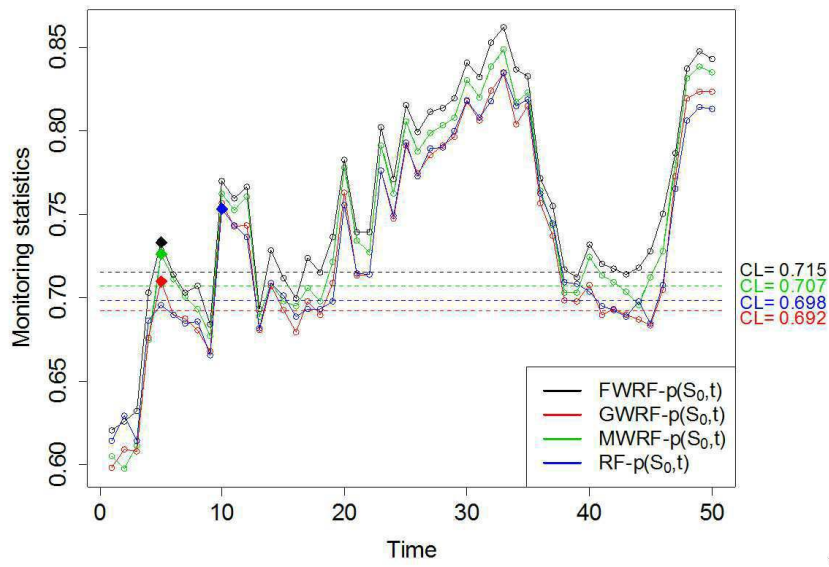
도면1



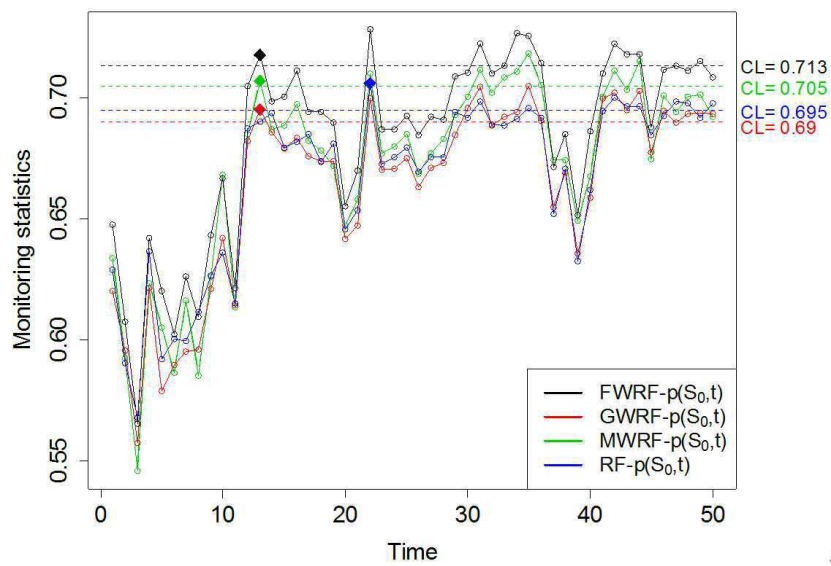
도면2



도면3



도면4



도면5

Comparison of ARL_1 of the proposed methods with the existing methods for 10-dimensional normal data.

Method	CL	ARL_0 (SDRL)	Scenario 1		Scenario 2	
			$\lambda = \sqrt{5}$	$\lambda = \sqrt{10}$	$\lambda = \sqrt{5}$	$\lambda = \sqrt{10}$
FWRP- $p(S_0, t)$	0.69	195.95 (8.63)	6.64 (0.18)	5.13* (0.14)	7.20* (0.16)	5.94 (0.12)
GWRF- $p(S_0, t)$	0.66	199.85 (8.88)	6.66 (0.17)	5.17 (0.14)	7.53 (0.17)	6.34 (0.12)
MWRP- $p(S_0, t)$	0.69	199.95 (8.93)	6.94 (0.18)	5.50 (0.14)	7.23 (0.16)	6.00 (0.12)
RF- $p(S_0, t)$	0.67	199.73 (8.90)	7.30 (0.18)	5.86 (0.14)	7.57 (0.18)	6.41 (0.13)
(Deng et al, 2012)						
D_0 -SVM	0.54	196.22 (8.72)	6.41* (0.57)	5.22 (0.28)	7.48 (0.65)	5.63* (0.42)
(He et al, 2016)						
LDA- D_1^2	3.93	199.47 (6.22)	5.03* (0.05)	2.79* (0.03)	6.05* (0.04)	4.30* (0.03)
(Wei et al, 2016)						

도면6

Comparison of ARL_1 of the proposed methods with the existing methods for 20-dimensional normal data.

Method	CL	ARL_0 (SDRL)	Scenario 1		Scenario 2	
			$\lambda = \sqrt{5}$	$\lambda = \sqrt{10}$	$\lambda = \sqrt{5}$	$\lambda = \sqrt{10}$
FWRP- $p(S_0, t)$	0.68	200.03 (8.99)	7.46* (0.22)	5.55 (0.11)	7.18* (0.15)	5.73 (0.10)
GWRF- $p(S_0, t)$	0.65	196.19 (8.72)	7.77 (0.22)	5.63 (0.10)	7.87 (0.17)	6.43 (0.10)
MWRP- $p(S_0, t)$	0.68	200.51 (8.88)	7.63 (0.23)	5.75 (0.12)	7.02 (0.16)	5.69* (0.11)
RF- $p(S_0, t)$	0.66	196.23 (8.65)	8.35 (0.21)	6.40 (0.14)	8.54 (0.19)	7.33 (0.13)
(Deng et al, 2012)						
D_0 -SVM	0.53	200.20 (9.24)	9.51 (0.49)	5.50* (0.12)	11.52 (0.73)	6.31 (0.18)
(He et al, 2016)						
LDA- D_1^2	7.78	200.03 (6.20)	6.27* (0.06)	4.33* (0.03)	6.11* (0.06)	4.15* (0.03)
(Wei et al, 2016)						

도면7

Comparison of ARL_1 of the proposed methods with the existing methods for 50-dimensional normal data.

Method	CL	ARL_0 (SDRL)	Scenario 1		Scenario 2	
			$\lambda = \sqrt{5}$	$\lambda = \sqrt{10}$	$\lambda = \sqrt{5}$	$\lambda = \sqrt{10}$
FWRP- $p(S_0, t)$	0.67	200.60 (8.72)	8.45 (0.41)	6.85 (0.15)	8.23 (0.20)	6.55 (0.10)
GWRF- $p(S_0, t)$	0.64	196.36 (8.21)	9.92 (0.55)	7.15 (0.17)	9.85 (0.31)	7.55 (0.11)
MWRP- $p(S_0, t)$	0.67	204.90 (9.02)	8.15* (0.38)	6.89 (0.17)	8.05* (0.19)	6.42* (0.10)
RF- $p(S_0, t)$	0.65	200.70 (8.84)	10.37 (0.34)	7.68 (0.17)	10.14 (0.28)	8.22 (0.14)
(Deng et al, 2012)						
D_0 -SVM	0.53	196.41 (8.89)	10.07 (0.38)	6.83* (0.15)	12.13 (0.56)	7.87 (0.20)
(He et al, 2016)						
LDA- D_1^2	24.00	199.97 (6.10)	7.79* (0.09)	5.77* (0.04)	8.40 (0.10)	5.43* (0.04)
(Wei et al, 2016)						

도면8

Comparison of ARL_1 of the proposed methods with the existing methods for 100-dimensional normal data.

Method	CL	ARL_0 (SDRL)	Scenario 1		Scenario 2	
			$\lambda = \sqrt{5}$	$\lambda = \sqrt{10}$	$\lambda = \sqrt{5}$	$\lambda = \sqrt{10}$
FWRF- $p(S_0, t)$	0.67	196.28 (8.87)	12.09 (0.67)	7.26 (0.23)	8.21 (0.21)	7.17 (0.08)
GWRF- $p(S_0, t)$	0.63	200.63 (8.98)	13.36 (0.82)	8.16 (0.28)	11.42 (0.43)	8.62 (0.11)
MWRF- $p(S_0, t)$	0.66	200.53 (8.93)	10.66* (0.54)	7.04* (0.20)	7.80*+ (0.18)	7.00* (0.09)
RF- $p(S_0, t)$ (Deng et al, 2012)	0.65	200.81 (8.88)	13.88 (0.54)	8.82 (0.19)	12.12 (0.38)	9.27 (0.16)
D_0 -SVM (He et al, 2016)	0.52	196.89 (8.78)	13.10 (0.56)	7.53 (0.15)	17.93 (0.97)	10.03 (0.30)
LDA- D_1^2 (Wei et al, 2016)	67.10	200.24 (6.10)	9.81+ (0.14)	6.37+ (0.04)	10.33 (0.14)	6.65+ (0.04)

도면9

Comparison of ARL_1 of the proposed methods with the existing methods for the bivariate t data.

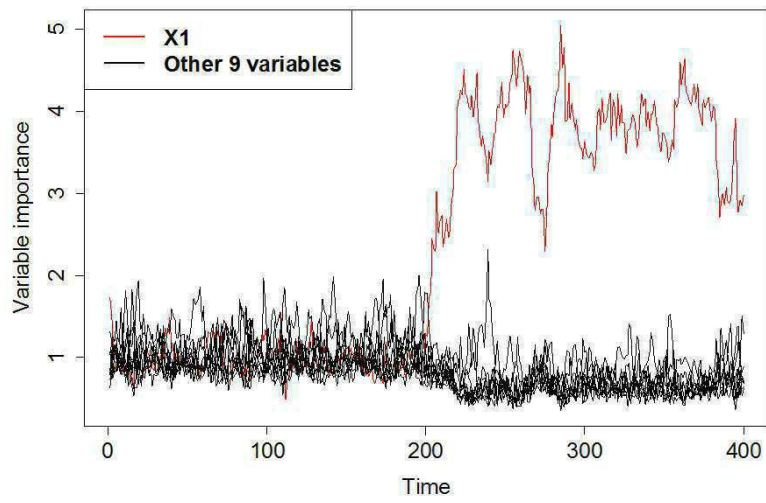
Method	CL	ARL_0 (SDRL)	Shift size (λ)					
			0.5	1.0	1.5	2.0	2.5	3.0
FWRF- $p(S_0, t)$	0.71	200.46 (8.85)	55.82 (2.97)	17.14 (0.66)	9.40 (0.22)	7.11 (0.13)	6.25 (0.10)	5.63 (0.09)
GWRF- $p(S_0, t)$	0.69	200.55 (8.80)	44.29+ (2.44)	14.99+ (0.54)	8.85+ (0.20)	6.99+ (0.13)	6.11+ (0.10)	5.53+ (0.09)
MWRF- $p(S_0, t)$	0.72	200.47 (8.63)	70.05 (3.62)	20.12 (0.83)	10.23 (0.26)	7.69 (0.13)	6.62 (0.10)	5.97 (0.09)
RF- $p(S_0, t)$ (Deng et al, 2012)	0.70	200.72 (8.85)	97.23 (4.88)	23.10 (0.98)	10.86 (0.29)	8.14 (0.15)	7.05 (0.11)	6.30 (0.09)
D_0 -SVM (He et al, 2016)	0.52	196.26 (8.79)	111.48 (6.06)	54.84 (2.98)	24.61 (1.17)	12.30 (0.60)	7.59 (0.25)	5.88 (0.14)
KLDA- D_0^2 (Wei et al, 2016)	7.28	199.98 (6.25)	147.66 (4.92)	51.44 (1.57)	17.65 (0.49)	8.68 (0.06)	6.50 (0.06)	5.86 (0.03)

도면10

Comparison of ARL_1 of the proposed methods with the existing methods for the bivariate gamma data.

Method	CL	ARL_0 (SDRL)	Shift size (λ)					
			0.5	1.0	1.5	2.0	2.5	3.0
FWRF- $p(S_0, t)$	0.70	200.48 (8.92)	15.92 (1.00)	7.23 (0.18)	6.03 (0.15)	5.33 (0.13)	4.84 (0.12)	4.59 (0.12)
GWRF- $p(S_0, t)$	0.69	204.42 (9.25)	12.93+ (0.73)	7.00+ (0.18)	5.88 (0.16)	5.25 (0.14)	4.80+ (0.13)	4.51+ (0.12)
MWRF- $p(S_0, t)$	0.71	204.45 (9.25)	19.51 (1.35)	7.57 (0.17)	6.32 (0.15)	5.62 (0.13)	5.19 (0.12)	4.92 (0.11)
RF- $p(S_0, t)$ (Deng et al, 2012)	0.69	200.38 (8.91)	20.96 (1.51)	7.83 (0.17)	6.58 (0.15)	5.95 (0.13)	5.48 (0.12)	5.18 (0.12)
D_0 -SVM (He et al, 2016)	0.51	205.06 (9.31)	39.23 (3.19)	12.11 (0.92)	6.75 (0.29)	5.64 (0.18)	5.10 (0.11)	4.90 (0.09)
KLDA- D_0^2 (Wei et al, 2016)	6.51	199.97 (6.30)	43.35 (1.29)	9.39 (0.17)	5.55+ (0.02)	5.00+ (0.01)	4.99 (0.01)	4.98 (0.01)

도면11



도면12

