



(19) 대한민국특허청(KR)

(12) 등록특허공보(B1)

(45) 공고일자 2021년06월17일

(11) 등록번호 10-2266182

(24) 등록일자 2021년06월11일

(51) 국제특허분류(Int. Cl.)
G06F 11/30 (2006.01) G06N 20/00 (2019.01)
G06Q 10/04 (2012.01)

(52) CPC특허분류
G06F 11/3089 (2013.01)
G06F 17/16 (2013.01)

(21) 출원번호 10-2020-0064892

(22) 출원일자 2020년05월29일

심사청구일자 2020년05월29일

(56) 선행기술조사문헌

KR1020180020626 A*

KR1020190129258 A*

KR1020180126353 A

KR1020190062739 A

*는 심사관에 의하여 인용된 문헌

(73) 특허권자

고려대학교 산학협력단

서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)

(72) 발명자

백준걸

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

홍재열

서울특별시 강북구 숭실로 174, 149동 1306호 (미아동, SK북한산시티아파트)

(74) 대리인

특허법인 신세기

전체 청구항 수 : 총 10 항

심사관 : 김계준

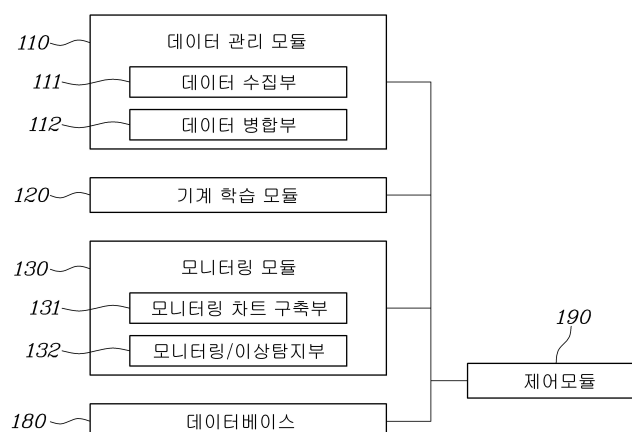
(54) 발명의 명칭 복수의 시계열 데이터의 모니터링 방법 및 장치

(57) 요약

k(k는 2 이상의 자연수)개의 파라미터에 각각에 대한 시계열 데이터를 수집하고, 시간 정보를 이용하여 상기 k개의 파라미터에 대한 시계열 데이터를 병합하여 프로파일 데이터를 생성하는 데이터 관리 모듈; 상기 프로파일 데이터를 입력 변수로 설정하고 상기 입력 변수에 따라 예측하고자 하는 예측치를 출력 변수로 설정하여 기계 학습을 수행하여 예측치를 획득하는 기계 학습 모듈; 및 상기 기계 학습 모듈에 의해 획득된 예측치를 기반으로 모니터링 대상 데이터에 대한 모니터링을 수행하며, 모니터링 결과 이상이 판단된 경우 이상이 발생한 시점의 예측치와 이상이 발생하지 않은 시점의 예측치에 대한 변수비율을 기반으로 상기 k 개의 파라미터 중 이상의 원인이 되는 파라미터를 판단하는 모니터링 모듈을 포함하는 복수의 시계열 데이터의 모니터링 장치가 개시된다.

대표도 - 도1

100



(52) CPC특허분류

G06N 20/00 (2019.01)

G06Q 10/04 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	2019R1A2C2005949
과제번호	R1904372
부처명	미래창조과학부
과제관리(전문)기관명	한국연구재단
연구사업명	중견연구자 지원
연구과제명	멀티 스트리밍 센서 데이터에 분석을 위한 캡슐화된 딥 네트워크 기법 개발
기 여 율	1/1
과제수행기관명	고려대학교 산학협력단
연구기간	2020.03.01 ~ 2021.02.28

명세서

청구범위

청구항 1

k(k는 2 이상의 자연수)개의 파라미터에 각각에 대한 시계열 데이터를 수집하고, 시간 정보를 이용하여 상기 k 개의 파라미터에 대한 시계열 데이터를 병합하여 프로파일 데이터를 생성하는 데이터 관리 모듈;

상기 프로파일 데이터를 입력 변수로 설정하고 상기 입력 변수에 따라 예측하고자 하는 예측치를 출력 변수로 설정하여 기계 학습을 수행하여 예측치를 획득하는 기계 학습 모듈; 및

상기 기계 학습 모듈에 의해 획득된 예측치를 기반으로 모니터링 대상 데이터에 대한 모니터링을 수행하며, 모니터링 결과 이상이 판단된 경우 이상이 발생한 시점의 예측치와 이상이 발생하지 않은 시점의 예측치에 대한 변수비율을 기반으로 상기 k 개의 파라미터 중 이상의 원인이 되는 파라미터를 판단하는 모니터링 모듈을 포함하며,

상기 예측치는 하기 식 1과 같이 표현되는 것을 특징으로 하는 복수의 시계열 데이터의 모니터링 장치.

[식 1]

$$\begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_t \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{t1} & x_{t2} & \dots & x_{tk} \end{bmatrix} \cdot \begin{bmatrix} \beta_0 \\ \beta_1 \\ \dots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_t \end{bmatrix}$$

('y_t'는 시점 't'에서의 예측치이고, 'x_{tk}'는 시점 't'에서의 각 파라미터의 시계열 데이터, ' β_k '는 k 개의 시계열 데이터 각각에 대한 기계학습 가중치이고, ' ε_t '는 시점 't'에서의 기계학습 바이어스값, ' β_k ' 및 ' ε_t '는 기계학습을 통해 결정됨).

청구항 2

청구항 1에 있어서, 상기 모니터링 모듈은,

상기 예측치를 중심선으로 하여 모니터링 차트를 구축하는 모니터링 차트 구축부; 및

상기 모니터링 차트를 이용하여 이상을 탐지하는 이상 탐지부를 포함하는 것을 특징으로 하는 복수의 시계열 데이터의 모니터링 장치.

청구항 3

청구항 2에 있어서,

상기 모니터링 차트 구축부는, 상기 중심선을 중심으로 평균이 0이고 표준편차가 σ 인 정규분포를 활용하여 상하방 관리한계선을 생성하며,

상기 이상 탐지부는 모니터링 대상 데이터가 상기 상하방 관리한계선을 사이의 영역을 벗어나면 이상이 발생한 것으로 판단하는 것을 특징으로 하는 복수의 시계열 데이터의 모니터링 장치.

청구항 4

삭제

청구항 5

청구항 1에 있어서, 상기 모니터링 모듈은,

상기 식 1을 이용하여 하기 식 2와 같이 각 시점에 대한 변수 비율을 정의하는 것을 특징으로 하는 복수의 시계

열 데이터의 모니터링 장치.

[식 2]

$$\text{변수비율} = \begin{bmatrix} \frac{x_{11} \cdot \beta_1}{y_1 - \varepsilon_1 - \beta_0} + \frac{x_{12} \cdot \beta_2}{y_2 - \varepsilon_2 - \beta_0} + \dots \frac{x_{1k} \cdot \beta_k}{y_t - \varepsilon_t - \beta_0} \\ \frac{x_{21} \cdot \beta_1}{y_1 - \varepsilon_1 - \beta_0} + \frac{x_{22} \cdot \beta_2}{y_2 - \varepsilon_2 - \beta_0} + \dots \frac{x_{2k} \cdot \beta_k}{y_t - \varepsilon_t - \beta_0} \\ \vdots \\ \frac{x_{t1} \cdot \beta_1}{y_1 - \varepsilon_1 - \beta_0} + \frac{x_{t2} \cdot \beta_2}{y_2 - \varepsilon_2 - \beta_0} + \dots \frac{x_{tk} \cdot \beta_k}{y_t - \varepsilon_t - \beta_0} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

(상기 식 2의 각 행은 특정 시점에서의 변수비율에 해당하며, 각 행을 구성하는 항은 상기 k 개의 파라미터 각각에 대한 변수 중요도에 해당함)

청구항 6

청구항 5에 있어서, 상기 모니터링 모듈은,

특정 시점의 모니터링 대상 데이터에 이상이 발생한 경우, 이상이 발생한 시점의 변수 비율과 이상이 발생하지 않은 변수 비율을 비교하며, 비교한 두 변수비율에서 특정 파라미터에 대한 변수 중요도의 차이가 사전 설정된 기준 이상 크게 나타나는 경우 해당 파라미터에 대한 시계열 데이터의 입력에 의해 이상이 발생하였음을 판단하는 것을 특징으로 하는 복수의 시계열 데이터의 모니터링 장치.

청구항 7

k(k는 2 이상의 자연수)개의 파라미터에 각각에 대한 시계열 데이터를 수집하는 단계;

시간 정보를 이용하여 상기 k개의 파라미터에 대한 시계열 데이터를 병합하여 프로파일 데이터를 생성하는 단계;

상기 프로파일 데이터를 입력 변수로 설정하고 상기 입력 변수에 따라 예측하고자 하는 예측치를 출력 변수로 설정하여 기계 학습을 수행하여 예측치를 획득하는 단계;

상기 예측치를 이용하여 모니터링 차트를 구축하는 단계; 및

상기 모니터링 차트를 이용하여 모니터링 대상 데이터의 모니터링을 수행하며, 모니터링 결과 이상이 판단된 경우 이상이 발생한 시점의 예측치와 이상이 발생하지 않은 시점의 예측치에 대한 변수비율을 기반으로 상기 k 개의 파라미터 중 이상의 원인이 되는 파라미터를 판단하는 단계를 포함하며,

상기 예측치는 하기 식 1과 같이 표현되는 것을 특징으로 하는 복수의 시계열 데이터의 모니터링 방법.

[식 1]

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_t \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{t1} & x_{t2} & \dots & x_{tk} \end{bmatrix} \cdot \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_t \end{bmatrix}$$

('y_t'는 시점 't'에서의 예측치이고, 'x_{tk}'는 시점 't'에서의 각 파라미터의 시계열 데이터, ' β_k '는 k 개의 시계열 데이터 각각에 대한 기계학습 가중치이고, ' ε_t '는 시점 't'에서의 기계학습 바이어스값 임. ' β_k ' 및 ' ε_t '는 기계학습을 통해 결정됨).

청구항 8

청구항 7에 있어서, 상기 모니터링 차트를 구축하는 단계는,

중심선을 생성하는 단계와, 상기 중심선을 기준으로 상하방 관리한계선을 생성하는 단계를 포함하며,

상기 판단하는 단계는,

모니터링 대상 데이터가 상기 상하방 관리한계선을 이탈하면 이상이 있는 것으로 탐지하고 모니터링 결과를 출

력하는 단계를 포함하는 것을 특징으로 하는 복수의 시계열 데이터의 모니터링 방법.

청구항 9

청구항 8에 있어서,

상기 중심선은 상기 기계 학습 단계를 통해 획득한 상기 예측치이며,

상기 관리한계선은 상기 중심선을 중심으로 평균이 0이고, 표준편차가 σ 인 정규분포를 활용하여 생성하는 것을 특징으로 하는 복수의 시계열 데이터의 모니터링 방법.

청구항 10

삭제

청구항 11

청구항 7에 있어서, 상기 모니터링을 수행하는 단계는,

상기 식 1을 이용하여 하기 식 2와 같이 각 시점에 대한 변수 비율을 정의하는 것을 특징으로 하는 복수의 시계열 데이터의 모니터링 방법.

[식 2]

$$\text{변수비율} = \begin{bmatrix} \frac{x_{11} \cdot \beta_1}{y_1 - \varepsilon_1 - \beta_0} + \frac{x_{12} \cdot \beta_2}{y_2 - \varepsilon_2 - \beta_0} + \dots \frac{x_{1k} \cdot \beta_k}{y_t - \varepsilon_t - \beta_0} \\ \frac{x_{21} \cdot \beta_1}{y_1 - \varepsilon_1 - \beta_0} + \frac{x_{22} \cdot \beta_2}{y_2 - \varepsilon_2 - \beta_0} + \dots \frac{x_{2k} \cdot \beta_k}{y_t - \varepsilon_t - \beta_0} \\ \vdots \\ \frac{x_{t1} \cdot \beta_1}{y_1 - \varepsilon_1 - \beta_0} + \frac{x_{t2} \cdot \beta_2}{y_2 - \varepsilon_2 - \beta_0} + \dots \frac{x_{tk} \cdot \beta_k}{y_t - \varepsilon_t - \beta_0} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

(상기 식 2의 각 행은 특정 시점에서의 변수비율에 해당하며, 각 행을 구성하는 항은 상기 k 개의 파라미터 각각에 대한 변수 중요도에 해당함)

청구항 12

청구항 11에 있어서, 상기 판단하는 단계는,

특정 시점의 모니터링 대상 데이터에 이상이 발생한 경우, 이상이 발생한 시점의 변수 비율과 이상이 발생하지 않은 변수 비율을 비교하며, 비교한 두 변수비율에서 특정 파라미터에 대한 변수 중요도의 차이가 사전 설정된 기준 이상 크게 나타나는 경우 해당 파라미터에 대한 시계열 데이터의 입력에 의해 이상이 발생하였음을 판단하는 것을 특징으로 하는 복수의 시계열 데이터의 모니터링 방법.

발명의 설명

기술 분야

[0001] 본 발명은 복수의 시계열 데이터의 모니터링 방법 및 장치에 관한 것으로, 더욱 상세하게는 복수의 시계열 데이터(스트리밍 데이터)를 하나의 다중 프로파일 데이터로 병합하고 기계학습 기법을 활용하여 모니터링 차트를 구축함으로써 복수의 시계열 데이터를 하나의 지표로 관리할 수 있는 복수의 시계열 데이터의 모니터링 방법 및 장치에 관한 것이다.

배경 기술

[0003] 시간의 흐름에 따라 값이 변경되는 시계열 데이터는 의료 분야에서는 심박수, 뇌파 등과 같은 신호에 해당하며, 금융 분야에서는 환율, 유가, 주가 등과 같은 데이터에 해당한다. 제조 분야도 마찬가지로 공정의 설비로부터 발생하는 여러 센서의 신호가 시계열 데이터로 나타난다. 이와 같이 다양한 분야에서 발생하는 시계열 데이터는 각 분야의 필요에 따라 모니터링 될 필요가 있다.

[0004] 종래의 시계열 데이터에 대한 모니터링 기법은 대부분 하나의 변수만 대상으로 하였다. 예를 들어, 금융 분야에

서 유가와 환율을 모니터링 하기 위해서는 각각 별도의 유가 모니터링 기법과 환율 모니터링 기법이 적용되었다.

[0005] 최근 모니터링 대상이 되는 시계열 데이터의 개수가 증가함에 따라, 종래와 같이 각각의 변수를 개별적으로 모니터링 하는 기법을 적용하게 되는 경우 다수의 시계열 데이터를 모니터링 하는 데 있어 성능이 저하되거나 비용이 증가하는 문제점이 있다. 따라서, 한번에 복수의 시계열 데이터를 모니터링 할 수 있는 새로운 모니터링 기법이 요구되고 있다. 나아가 복수의 시계열 데이터 모니터링 시 이상이 발생된 경우 이상에 대한 분석까지 가능한 모니터링 기법이 요구되고 있다.

[0006]

[0007] 이상에서 배경기술로서 설명된 사항들은 본 발명의 배경에 대한 이해 증진을 위한 것일 뿐, 이 기술분야에서 통상의 지식을 가진 자에게 이미 알려진 종래기술에 해당함을 인정하는 것으로 받아들여져서는 안 될 것이다.

발명의 내용

해결하려는 과제

[0009] 이에 본 발명은, 복수의 시계열 데이터를 하나의 다중 프로파일 데이터로 병합하고 기계학습 기법을 활용하여 모니터링 차트를 구축함으로써 복수의 시계열 데이터를 하나의 지표로 관리할 수 있는 복수의 시계열 데이터의 모니터링 방법 및 장치를 제공하는 것을 해결하고자 하는 기술적 과제로 한다.

[0010] 나아가, 본 발명은 하나의 지표로 관리 모니터링 되는 복수의 시계열 데이터에서 이상이 발생한 경우 이상이 발생한 데이터를 판별할 수 있는 이상 분석이 가능한 복수의 시계열 데이터의 모니터링 방법 및 장치를 제공하는 것을 해결하고자 하는 기술적 과제로 한다.

과제의 해결 수단

[0012] 상기 기술적 과제를 해결하기 위한 수단으로서 본 발명은,

[0013] k(k는 2 이상의 자연수)개의 파라미터에 각각에 대한 시계열 데이터를 수집하고, 시간 정보를 이용하여 상기 k 개의 파라미터에 대한 시계열 데이터를 병합하여 프로파일 데이터를 생성하는 데이터 관리 모듈;

[0014] 상기 프로파일 데이터를 입력 변수로 설정하고 상기 입력 변수에 따라 예측하고자 하는 예측치를 출력 변수로 설정하여 기계 학습을 수행하여 예측치를 획득하는 기계 학습 모듈; 및

[0015] 상기 기계 학습 모듈에 의해 획득된 예측치를 기반으로 모니터링 대상 데이터에 대한 모니터링을 수행하며, 모니터링 결과 이상이 판단된 경우 이상이 발생한 시점의 예측치와 이상이 발생하지 않은 시점의 예측치에 대한 변수비율을 기반으로 상기 k 개의 파라미터 중 이상의 원인이 되는 파라미터를 판단하는 모니터링 모듈;

[0016] 을 포함하는 복수의 시계열 데이터의 모니터링 장치가 개시된다.

[0017] 본 발명의 일 실시형태에서, 상기 모니터링 모듈은, 상기 예측치를 중심선으로 하여 모니터링 차트를 구축하는 모니터링 차트 구축부; 및 상기 모니터링 차트를 이용하여 이상을 탐지하는 이상 탐지부를 포함할 수 있다.

[0018] 본 발명의 일 실시형태에서, 상기 모니터링 차트 구축부는, 상기 중심선을 중심으로 평균이 0이고 표준편차가 σ 인 정규분포를 활용하여 상하방 관리한계선을 생성하며, 상기 이상 탐지부는 모니터링 대상 데이터가 상기 상하방 관리한계선을 사이의 영역을 벗어나면 이상이 발생한 것으로 판단할 수 있다.

[0019] 본 발명의 일 실시형태에서, 상기 예측치는 하기 식 1과 같이 표현될 수 있다.

[0020] [식 1]

$$\begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_t \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{t1} & x_{t2} & \dots & x_{tk} \end{bmatrix} \cdot \begin{bmatrix} \beta_0 \\ \beta_1 \\ \dots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_t \end{bmatrix}$$

[0022] 상기 식 1에서 ' y_t '는 시점 ' t '에서의 예측치이고, ' x_{tk} '는 시점 ' t '에서의 각 파라미터의 시계열 데이터, ' β_k '는 k 개의 시계열 데이터 각각에 대한 기계학습 가중치이고, ' ε_t '는 시점 ' t '에서의 기계학습 바이어스값이다. '

β_k 및 ϵ_t 는 기계학습을 통해 결정될 수 있다.

[0023] 본 발명의 일 실시형태에서, 상기 모니터링 모듈은, 상기 식 1을 이용하여 하기 식 2와 같이 각 시점에 대한 변수 비율을 정의할 수 있다.

[0024] [식 2]

$$\text{변수비율} = \begin{bmatrix} \frac{x_{11} \cdot \beta_1}{y_1 - \epsilon_1 - \beta_0} + \frac{x_{12} \cdot \beta_2}{y_2 - \epsilon_2 - \beta_0} + \dots \frac{x_{1k} \cdot \beta_k}{y_t - \epsilon_t - \beta_0} \\ \frac{x_{21} \cdot \beta_1}{y_1 - \epsilon_1 - \beta_0} + \frac{x_{22} \cdot \beta_2}{y_2 - \epsilon_2 - \beta_0} + \dots \frac{x_{2k} \cdot \beta_k}{y_t - \epsilon_t - \beta_0} \\ \dots \\ \frac{x_{t1} \cdot \beta_1}{y_1 - \epsilon_1 - \beta_0} + \frac{x_{t2} \cdot \beta_2}{y_2 - \epsilon_2 - \beta_0} + \dots \frac{x_{tk} \cdot \beta_k}{y_t - \epsilon_t - \beta_0} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \dots \\ 1 \end{bmatrix}$$

[0025]

[0026] 상기 식 2의 각 행은 특정 시점에서의 변수비율에 해당하며, 각 행을 구성하는 항은 상기 k 개의 파라미터 각각에 대한 변수 중요도에 해당한다.

[0027] 본 발명의 일 실시형태에서, 상기 모니터링 모듈은, 특정 시점의 모니터링 대상 데이터에 이상이 발생한 경우, 이상이 발생한 시점의 변수 비율과 이상이 발생하지 않은 변수 비율을 비교하며, 비교한 두 변수비율에서 특정 파라미터에 대한 변수 중요도의 차이가 사전 설정된 기준 이상 크게 나타나는 경우 해당 파라미터에 대한 시계열 데이터의 입력에 의해 이상이 발생하였음을 판단할 수 있다.

[0028] 상기 기술적 과제를 해결하기 위한 다른 수단으로서 본 발명은,

[0029] k(k는 2 이상의 자연수)개의 파라미터에 각각에 대한 시계열 데이터를 수집하는 단계;

[0030] 시간 정보를 이용하여 상기 k개의 파라미터에 대한 시계열 데이터를 병합하여 프로파일 데이터를 생성하는 단계;

[0031] 상기 프로파일 데이터를 입력 변수로 설정하고 상기 입력 변수에 따라 예측하고자 하는 예측치를 출력 변수로 설정하여 기계 학습을 수행하여 예측치를 획득하는 단계;

[0032] 상기 예측치를 이용하여 모니터링 차트를 구축하는 단계; 및

[0033] 상기 모니터링 차트를 이용하여 모니터링 대상 데이터의 모니터링을 수행하며, 모니터링 결과 이상이 판단된 경우 이상이 발생한 시점의 예측치와 이상이 발생하지 않은 시점의 예측치에 대한 변수비율을 기반으로 상기 k 개의 파라미터 중 이상의 원인이 되는 파라미터를 판단하는 단계;

[0034] 를 포함하는 복수의 시계열 데이터의 모니터링 방법을 제공한다.

[0035] 본 발명의 일 실시형태에서, 상기 모니터링 차트를 구축하는 단계는, 중심선을 생성하는 단계와, 상기 중심선을 기준으로 상하방 관리한계선을 생성하는 단계를 포함하며, 상기 판단하는 단계는, 모니터링 대상 데이터가 상기 상하방 관리한계선을 이탈하면 이상이 있는 것으로 탐지하고 모니터링 결과를 출력하는 단계를 포함할 수 있다.

[0036] 본 발명의 일 실시형태에서, 상기 중심선은 상기 기계 학습 단계를 통해 획득한 상기 예측치이며, 상기 관리한계선은 상기 중심선을 중심으로 평균이 0이고, 표준편차가 σ 인 정규분포를 활용하여 생성할 수 있다.

[0037] 본 발명의 일 실시형태에서, 상기 예측치는 하기 식 3과 같이 표현될 수 있다.

[0038] [식 3]

$$\begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_t \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{t1} & x_{t2} & \dots & x_{tk} \end{bmatrix} \cdot \begin{bmatrix} \beta_0 \\ \beta_1 \\ \dots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \dots \\ \epsilon_t \end{bmatrix}$$

[0039]

[0040] 식 3에서, ' y_t '는 시점 ' t '에서의 예측치이고, ' x_{tk} '는 시점 ' t '에서의 각 파라미터의 시계열 데이터, ' β_k '는 k 개의 시계열 데이터 각각에 대한 기계학습 가중치이고, ' ϵ_t '는 시점 ' t '에서의 기계학습 바이어스값이다. β_k

및 ' ε_t '는 기계학습을 통해 결정될 수 있다.

[0041] 본 발명의 일 실시형태에서, 상기 모니터링을 수행하는 단계는, 상기 식 4를 이용하여 하기 식 2와 같이 각 시점에 대한 변수 비율을 정의하는 것을 특징으로 하는 복수의 시계열 데이터의 모니터링 방법.

[0042] [식 4]

$$\text{변수비율} = \begin{bmatrix} \frac{x_{11} \cdot \beta_1}{y_1 - \varepsilon_1 - \beta_0} + \frac{x_{12} \cdot \beta_2}{y_2 - \varepsilon_2 - \beta_0} + \dots \frac{x_{1k} \cdot \beta_k}{y_t - \varepsilon_t - \beta_0} \\ \frac{x_{21} \cdot \beta_1}{y_1 - \varepsilon_1 - \beta_0} + \frac{x_{22} \cdot \beta_2}{y_2 - \varepsilon_2 - \beta_0} + \dots \frac{x_{2k} \cdot \beta_k}{y_t - \varepsilon_t - \beta_0} \\ \vdots \\ \frac{x_{t1} \cdot \beta_1}{y_1 - \varepsilon_1 - \beta_0} + \frac{x_{t2} \cdot \beta_2}{y_2 - \varepsilon_2 - \beta_0} + \dots \frac{x_{tk} \cdot \beta_k}{y_t - \varepsilon_t - \beta_0} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

[0043]

[0044] 식 4의 각 행은 특정 시점에서의 변수비율에 해당하며, 각 행을 구성하는 항은 상기 k 개의 파라미터 각각에 대한 변수 중요도에 해당할 수 있다.

[0045] 본 발명의 일 실시형태에서, 상기 판단하는 단계는, 특정 시점의 모니터링 대상 데이터에 이상이 발생한 경우, 이상이 발생한 시점의 변수 비율과 이상이 발생하지 않은 변수 비율을 비교하며, 비교한 두 변수비율에서 특정 파라미터에 대한 변수 중요도의 차이가 사전 설정된 기준 이상 크게 나타나는 경우 해당 파라미터에 대한 시계열 데이터의 입력에 의해 이상이 발생하였음을 판단할 수 있다.

발명의 효과

[0047] 상기 복수의 시계열 데이터의 모니터링 방법 및 장치에 따르면, 복수의 시계열 데이터를 하나의 프로파일 데이터로 전처리하고 기계 학습을 수행하여 얻어진 정보들을 이용하여 모니터링 차트를 구축함으로써, 복수의 파라미터에 대한 데이터를 하나의 지표로 관리하고, 데이터의 이상 여부를 모니터링할 수 있다.

[0048] 특히, 상기 복수의 시계열 데이터의 모니터링 방법 및 장치에 따르면, 데이터의 이상이 발생한 경우 각 시점에서 정의된 변수 비율을 이용하여 변수 비율에서 특정 시계열 데이터가 차지하는 변수 중요도를 판단함으로써 이상이 발생한 시점에서 이상의 원인이 되는 변수를 정확하게 판별할 수 있다.

[0050] 본 발명에서 얻을 수 있는 효과는 이상에서 언급한 효과들로 제한되지 않으며, 언급하지 않은 또 다른 효과들은 아래의 기재로부터 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에게 명확하게 이해될 수 있을 것이다.

도면의 간단한 설명

[0052] 도 1은 본 발명의 일 실시형태에 따른 복수의 시계열 데이터 모니터링 장치를 도시한 블록 구성도이다.

도 2는 본 발명의 일 실시형태에 따른 복수의 시계열 데이터의 모니터링 방법을 도시한 흐름도이다.

도 3은 본 발명의 일 실시형태에 따른 복수의 시계열 데이터의 모니터링 방법 및 장치에서 사용되는 복수의 시계열 데이터의 예를 도시한 그래프이다.

도 4는 도 3에 도시된 복수의 시계열 데이터를 이용한 기계학습에 의해 도출된 예측치를 사용하여 구축된 모니터링 차트의 일예를 도시한 도면이다.

도 5는 도 4에 도시된 모니터링 차트를 통해 이상 발생 여부 판단 시 이상이 발생한 경우 이상 발생 원인을 분석하는 기법을 설명하는 도면이다.

발명을 실시하기 위한 구체적인 내용

[0053] 이하, 첨부 도면을 참조하여 다양한 실시 형태에 따른 복수의 시계열 데이터의 모니터링 방법 및 장치를 상세하게 설명하기로 한다.

[0054] 도 1은 본 발명의 일 실시형태에 따른 복수의 시계열 데이터 모니터링 장치를 도시한 블록 구성도이다.

[0055] 도 1을 참조하면, 본 발명의 일 실시형태에 따른 복수의 시계열 데이터 모니터링 장치(100)는 복수의 시계열 데이터를 수집하고 병합하여 기계 학습을 위한 전처리를 수행하는 데이터 관리 모듈(110), 기계 학습 모듈(120),

모니터링 차트를 구축하여 모니터링을 수행하며 이상을 탐지하는 모니터링부(130), 데이터베이스(180), 및 제어 모듈(190)을 포함할 수 있다.

[0056] 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.

[0057] 본 발명의 일 실시형태에 따른 복수의 시계열 데이터 모니터링 장치(100) 내 데이터 관리 모듈(110)은 데이터 수집부(111) 및 데이터 병합부(112)를 포함할 수 있다.

[0058] 데이터 수집부(111)는 k (k 는 2 이상의 자연수) 개의 파라미터에 대한 각각의 시계열 데이터를 수집하고 이를 데이터베이스(180)에 저장할 수 있다. 이 때, 시계열 데이터와 관련된 시간 정보를 함께 수집하여 저장한다. 예를 들어, 시계열 데이터가 제조 공정의 이상 탐지를 위한 복수의 센서로부터 수신된 센서 데이터인 경우, 데이터 수집부(111)는 복수의 센서로부터 수신된 k (k 는 2 이상의 자연수) 개의 파라미터를 갖는 센서 데이터(센서 값) 및 관련 시간 정보(예를 들어, 센싱 시간 등)를 수집하여 데이터베이스(180)에 저장할 수 있다.

[0059] 데이터 병합부(112)는 데이터베이스(180)에 저장된 데이터와 시간 정보를 활용하여 k (k 는 2 이상의 자연수) 개의 파라미터에 대한 각각의 데이터를 하나의 프로파일 데이터로 병합할 수 있다. 데이터 병합은 기계학습 알고리즘을 사용하기 위해 거치는 전처리 단계로서 복수의 파라미터에 대한 각각의 데이터를 하나의 프로파일 형태의 데이터로 재가공 할 수 있다. 예를 들어, 시계열 데이터가 제조 공정의 이상 탐지를 위한 복수의 센서로부터 수신된 센서 데이터인 경우, 공정 시간(Process Time)에 따라 데이터를 병합할 수 있다.

[0060] 본 발명의 일 실시형태에 따른 복수의 시계열 데이터 모니터링 장치(100) 내 기계 학습 모듈(120)은 데이터 관리 모듈(110)에서 재가공된 프로파일 데이터를 이용하여 기계 학습을 수행한다. 즉, 데이터 관리 모듈(110)에서 가공된 다중 프로파일 데이터를 입력 변수로 설정하고, 상기 입력 변수에 따라 획득하고자 하는 예측치(y_1 - y_t)를 출력 변수로 설정하여 기계 학습을 수행하여 예측치를 획득하고, 획득한 예측치를 데이터베이스(180)에 저장한다.

[0061] 기계 학습을 통해 획득되는 예측치는 다음의 식 1과 같이 표현될 수 있다.

[0062] [식 1]

$$\begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_t \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{t1} & x_{t2} & \dots & x_{tk} \end{bmatrix} \cdot \begin{bmatrix} \beta_0 \\ \beta_1 \\ \dots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_t \end{bmatrix}$$

[0063] [0064] 상기 식 1은 총 k (k 는 2 이상의 자연수) 개의 파라미터에 대한 시계열 데이터에 대해 적용된 것으로, ' y_t '는 시점 ' t '에서의 예측치이고, ' x_{tk} '는 시점 ' t '에서의 각 파라미터의 시계열 데이터에 해당한다. ' β_k '는 k 개의 시계열 데이터 각각에 대한 기계학습 가중치이고, ' ε_t '는 시점 ' t '에서의 기계학습 바이어스값에 해당한다. ' β_k '

나 ' ε_t '는 기계학습을 통해 결정될 수 있다. 상기 식 1에서 행렬 $\begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_t \end{bmatrix}$ 는 기계학습의 출력 변수가 되고 행렬

$\begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{t1} & x_{t2} & \dots & x_{tk} \end{bmatrix}$ 는 병합된 복수의 시계열 데이터로서 기계학습의 입력 변수가 될 수 있다.

[0065] 본 발명의 일 실시형태에 따른 복수의 시계열 데이터 모니터링 장치(100) 내 모니터링 모듈(130)은 모니터링 차트 구축부(131) 및 모니터링/이상탐지부(132)를 포함할 수 있다.

[0066] 모니터링 차트 구축부(131)는 기계 학습을 통해 식 1과 같이 추출된 프로파일 데이터의 예측치를 활용하여 모니터링 차트를 구축할 수 있다. 모니터링 차트 구축 방법에 대해서는 후술하는 복수의 시계열 데이터의 모니터링 방법에서 더욱 상세하게 설명될 것이다.

- [0067] 모니터링/이상탐지부(132)는 구축된 모니터링 차트를 이용하여 모니터링을 수행하고 이상을 탐지한다. 즉, 모니터링을 통해 모니터링 차트의 관리한계선 범위 내에 속하는 경우 정상으로, 관리한계선을 벗어나는 경우 이상으로 판단하고, 이상/정상 탐지 결과를 출력한다. 모니터링 결과는 데이터베이스에 저장할 수 있다.
- [0068] 더하여, 모니터링/이상탐지부(132)는 관리한계선을 벗어나는 이상이 발생하는 경우 상기 식 1을 통해 도출되는 변수 중요도를 기반으로 이상이 발생한 시계열 데이터를 판단할 수 있다. 모니터링/이상탐지부(132)에 의해 수행되는 이상 판단 기법에 대해서도 후술하는 복수의 시계열 데이터의 모니터링 방법에서 더욱 상세하게 설명될 것이다.
- [0069] 데이터베이스(180)는 복수의 파라미터에 대하여 수집된 시계열 데이터 및 데이터 관련 시간 정보를 저장할 수 있다. 또한, 복수의 파라미터에 대한 데이터를 시간 정보를 활용하여 병합한 프로파일 데이터가 저장될 수 있다. 또한, 데이터베이스(180)에는 기계 학습 결과, 이를 통하여 구축된 모니터링 차트, 및 모니터링 결과 등이 저장될 수 있다. 본 명세서에서 데이터베이스라 함은, 각각의 데이터베이스에 대응되는 정보를 저장하는 소프트웨어 및 하드웨어의 기능적 구조적 결합을 의미할 수도 있다.
- [0070] 제어모듈(190)은 본 발명의 여러 실시형태에 따른 복수의 시계열 데이터의 모니터링 장치(100)의 전반적인 동작을 제어하는 요소이다. 즉, 제어모듈(190)은 데이터 처리 모듈(110), 기계 학습 모듈(120), 모니터링 모듈(130), 및 데이터베이스(180)의 동작을 제어할 수 있다.
- [0071] 이하, 본 발명의 일 실시형태에 따른 복수의 시계열 데이터의 모니터링 방법을 설명한다. 이하에서 이루어지는 본 발명의 일 실시형태에 따른 복수의 시계열 데이터의 모니터링 방법에 대한 설명을 통해 전술한 본 발명의 일 실시형태에 따른 복수의 시계열 데이터의 모니터링 장치의 동작이 더욱 명확하게 이해될 수 있을 것이다.
- [0072] 도 2는 본 발명의 일 실시형태에 따른 복수의 시계열 데이터의 모니터링 방법을 도시한 흐름도이다.
- [0073] 도 2를 참조하면, 본 발명의 일 실시형태에 따른 복수의 시계열 데이터의 모니터링 방법은, 복수의 파라미터 각각에 대한 복수의 시계열 데이터를 수집하고 수집된 복수의 시계열 데이터를 하나의 프로파일 데이터로 병합하는 데이터 처리 단계(S100), 기계 학습을 수행하여 예측치를 획득하는 기계 학습 단계(S200), 및 모니터링 차트를 구축하고 모니터링을 수행하여 이상 여부를 탐지하는 모니터링 단계(S300)를 포함할 수 있다.
- [0074] 먼저, 데이터 처리 단계(S100)에서, 데이터 관리 모듈(110)은 k (k 는 2 이상의 자연수) 개의 파라미터에 대한 각각의 데이터를 수신하고(S110) 이를 데이터베이스에 저장할 수 있다. 이 때, 데이터와 관련된 시간 정보가 함께 수신되어 저장될 수 있다. 예를 들어, 복수의 시계열 데이터가 제조 공정의 이상 탐지를 위한 복수의 센서로부터 각각 수신된 센서 데이터인 경우, 데이터 관리 모듈(110)은, 복수의 센서로부터 수신된 k (k 는 2 이상의 자연수) 개의 종류를 갖는 센서 데이터 및 관련 시간 정보(예를 들어, 센싱 시간 등)를 수집하여 데이터베이스에 저장할 수 있다.
- [0075] 이어, 데이터 처리 단계(S100)에서, 데이터 관리 모듈(110)은 복수의 시계열 데이터를 병합하는 단계(S120)를 수행할 수 있다. 구체적으로, 단계(S120)에서 데이터 관리 모듈(110)은 데이터베이스에 저장된 데이터와 시간 정보를 활용하여 복수의 시계열 데이터를 하나의 프로파일 데이터 형태로 병합한다. 데이터 병합은 기계학습 알고리즘을 사용하기 위해 거치는 전처리 단계로서 복잡한 데이터들을 하나의 프로파일 형태의 데이터로 재가공한다.
- [0076] 도 3은 본 발명의 일 실시형태에 따른 복수의 시계열 데이터의 모니터링 방법 및 장치에서 사용되는 복수의 시계열 데이터의 예를 도시한 그래프이다. 특히 도 3은 복수의 시계열 데이터가 제조 공정의 이상 탐지를 위한 복수의 센서로부터 각각 수신된 센서 데이터인 경우를 예로 도시한 것으로, 총 k (k 는 2 이상의 자연수) 개의 센서로부터 입력되는 시계열 데이터를 나타낸다.
- [0077] 도 3을 참조하면, 각 센서로부터 입력되는 복수의 시계열 데이터는 동일한 시간대에 발생하기 때문에 공정 시간(Process Time)에 따라 데이터를 병합할 수 있다. 예를 들면, t 시점의 데이터는 병합 후에는 데이터 테이블의 한 행($x_{t1}, x_{t2}, \dots, x_{tk}$)으로 표시될 수 있다. 이처럼 공정 시간이 진행되면서 데이터 테이블은 계속 병합되고 누적되어 기계학습 알고리즘에 적용하기 위한 전처리가 완료된다. 이때, t 는 공정 시간을 의미하고, x_{tk} 는 각 공
- 정 시점에 대한 제 k 센서 데이터 측정치를 나타낸다. 상기 식 1에서 우변의 행렬
- $$\begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & x_{t1} & x_{t2} & \cdots & x_{tk} \end{bmatrix}$$

같이 병합된 복수의 시계열 데이터가 표현될 수 있다.

- [0078] 이어, 본 발명의 일 실시형태에 따른 복수의 시계열 데이터의 모니터링 장치(100)의 기계 학습 모듈(120)은 데이터 관리 모듈(110)에서 재가공된 프로파일 데이터를 이용하여 기계 학습을 수행할 수 있다(S200).
- [0079] 기계 학습 수행 단계(S200)에서는, 데이터 처리 단계(S100)에서 가공된 복수의 시계열 데이터의 병합치를 입력 변수로 설정하고, 예측치(y_t)를 출력 변수로 설정하여 기계 학습을 수행할 수 있다(S210). 이러한 기계 학습을 통해 예측치가 획득될 수 있으며(S220), 획득된 예측치는 데이터베이스에 저장될 수 있다. 기계 학습을 통해 획득된 예측치는 상기 식 1과 같은 형태로 표현될 수 있다.
- [0080] 이어, 본 발명의 일 실시형태에 따른 복수의 시계열 데이터의 모니터링 장치(100)의 모니터링 모듈(130)은 기계 학습을 통해 추출된 예측치를 활용하여 모니터링 차트를 구축할 수 있다(S310).
- [0081] 도 4는 도 3에 도시된 복수의 시계열 데이터를 이용한 기계학습에 의해 도출된 예측치를 사용하여 구축된 모니터링 차트의 일예를 도시한 도면이다.
- [0082] 도 4를 참조하면, 모니터링 차트 구축 단계(S310)에서 예측치의 중심선(CL: Center Line) 및 상하방 관리한계선(UCL: Upper Control Limits, LCL: Lower Control Limit)이 산출될 수 있다.
- [0083] 중심선(CL)은 기계학습 알고리즘을 통해 계산된 출력변수 예측치($\hat{y}$)를 사용하여 다음의 식 2와 같이 표현될 수 있으며, 정상 상태를 결정하는 상하방 관리한계선(UCL, LCL)은 다음의 식 3 및 식 4와 같이 표현될 수 있다.
- [0084] [식 2]
- [0085]
$$CL = \hat{y}$$
- [0086] [식 3]
- [0087]
$$UCL = \hat{y} + C.V * \sigma$$
- [0088] [식 4]
- [0089]
$$LCL = \hat{y} - C.V * \sigma$$
- [0090] 상기 식 3 및 식 4에 표현된 바와 같이, 상하방 관리한계선(UCL, LCL)은 중심선(CL)을 중심으로 평균이 0이고 표준편차가 σ 인 정규분포를 활용하여 결정될 수 있다.
- [0091] 상기 식 3 및 식 4에서 C.V(Critical Value)는 관리한계선의 구간을 결정하는 결정계수로서, 제 현장의 상황 또는 모니터링 대상에 맞게 관리자 또는 사용자 등에 의하여 조절 가능하다.
- [0092] 이어, 본 발명의 일 실시형태에 따른 복수의 시계열 데이터의 모니터링 장치(100)의 모니터링 모듈(130)은 구축된 모니터링 차트를 이용하여 모니터링 단계(S320)를 수행할 수 있다.
- [0093] 이어, 모니터링 모듈(130)은 모니터링 단계(S320)에서 수행되는 모니터링을 통해, 모니터링 대상 데이터가 모니터링 차트의 관리한계선을 벗어나는지를 판단하는 이상 탐지 단계를 수행할 수 있다(S330). 모니터링 대상 데이터는 실제 입력되는 복수의 시계열 데이터를 병합한 프로파일 데이터가 될 수 있다. 이상이 탐지된 경우, 모니터링 모듈(130)은 이상 발생의 원인을 분석하고 이상 탐지 정보 및 이상 발생 원인을 출력할 수 있다(S340). 모니터링 결과들은 데이터베이스에 저장될 수 있다.
- [0094] 더욱 구체적으로, 이상 발생 원인 분석 및 출력 단계(S340)에서, 모니터링 모듈(130)은 모니터링 모듈(130)은 이상이 탐지될 경우, 도 5에 도시된 것과 같이 원인 분석을 실시할 수 있다.
- [0095] 도 5는 도 4에 도시된 모니터링 차트를 통해 이상 발생 여부 판단 시 이상이 발생한 경우 이상 발생 원인을 분석하는 기법을 설명하는 도면이다.
- [0096] 도 5를 참조하면, 예측치가 상하방 관리한계선을 벗어나는 이상 상태가 발생한 경우, 이상 상태를 나타내는 시점(t_1)의 변수 중요도와 이상 상태가 나타나지 않는 시점 중 하나(t_2)의 변수 중요도를 비교하여 어떠한 입력 시계열 데이터에서 이상이 발생하였는지 판단할 수 있다.
- [0097] 상기 식 1을 다음의 식 5의 과정을 통해 정리하면 식 6과 같이 변수 비율(Proportion of variable)을 도출할 수

있으며, 이 변수 비율을 구하는 식의 각 항을 변수 중요도로 정의할 수 있다.

[식 5]

$$\begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_t \end{bmatrix} - \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_t \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{t1} & x_{t2} & \dots & x_{tk} \end{bmatrix} \cdot \begin{bmatrix} \beta_0 \\ \beta_1 \\ \dots \\ \beta_k \end{bmatrix}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_t \end{bmatrix} - \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_t \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{t1} & x_{t2} & \dots & x_{tk} \end{bmatrix} \cdot \begin{bmatrix} \beta_0 \\ \beta_1 \\ \dots \\ \beta_k \end{bmatrix} = \begin{bmatrix} \beta_0 \\ \beta_0 \\ \dots \\ \beta_0 \end{bmatrix} + \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1k} \\ x_{21} & x_{22} & \dots & x_{2k} \\ \dots & \dots & \dots & \dots \\ x_{t1} & x_{t2} & \dots & x_{tk} \end{bmatrix} \cdot \begin{bmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_k \end{bmatrix}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_t \end{bmatrix} - \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_t \end{bmatrix} - \begin{bmatrix} \beta_0 \\ \beta_0 \\ \dots \\ \beta_0 \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1k} \\ x_{21} & x_{22} & \dots & x_{2k} \\ \dots & \dots & \dots & \dots \\ x_{t1} & x_{t2} & \dots & x_{tk} \end{bmatrix} \cdot \begin{bmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_k \end{bmatrix}$$

$$\begin{bmatrix} y_1 - \varepsilon_1 - \beta_0 \\ y_2 - \varepsilon_2 - \beta_0 \\ \dots \\ y_t - \varepsilon_t - \beta_0 \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1k} \\ x_{21} & x_{22} & \dots & x_{2k} \\ \dots & \dots & \dots & \dots \\ x_{t1} & x_{t2} & \dots & x_{tk} \end{bmatrix} \cdot \begin{bmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_k \end{bmatrix}$$

$$\begin{bmatrix} x_{11} & x_{12} & x_{1k} \\ x_{21} & x_{22} & x_{2k} \\ \dots & \dots & \dots \\ x_{t1} & x_{t2} & x_{tk} \end{bmatrix} \cdot \begin{bmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_k \end{bmatrix} / \begin{bmatrix} y_1 - \varepsilon_1 - \beta_0 \\ y_2 - \varepsilon_2 - \beta_0 \\ \dots \\ y_t - \varepsilon_t - \beta_0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \dots \\ 1 \end{bmatrix}$$

[식 6]

$$\text{Proportion of variable} = \begin{bmatrix} \frac{x_{11} \cdot \beta_1}{y_1 - \varepsilon_1 - \beta_0} + \frac{x_{12} \cdot \beta_2}{y_2 - \varepsilon_2 - \beta_0} + \dots + \frac{x_{1k} \cdot \beta_k}{y_t - \varepsilon_t - \beta_0} \\ \frac{x_{21} \cdot \beta_1}{y_1 - \varepsilon_1 - \beta_0} + \frac{x_{22} \cdot \beta_2}{y_2 - \varepsilon_2 - \beta_0} + \dots + \frac{x_{2k} \cdot \beta_k}{y_t - \varepsilon_t - \beta_0} \\ \dots \\ \frac{x_{t1} \cdot \beta_1}{y_1 - \varepsilon_1 - \beta_0} + \frac{x_{t2} \cdot \beta_2}{y_2 - \varepsilon_2 - \beta_0} + \dots + \frac{x_{tk} \cdot \beta_k}{y_t - \varepsilon_t - \beta_0} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \dots \\ 1 \end{bmatrix}$$

상기 식 6에 나타난 바와 같이, 각 시점별 변수 비율의 크기는 1이 되며, 각 시점별 변수비율을 나타내는 각 항들이 입력된 파라미터에 대한 변수 중요도가 된다. 변수 중요도는 특정 시점에 입력된 각 시계열 데이터가 병합된 하나의 프로파일에서 차지하는 크기를 나타낸 것이다.

도 5의 예에서 나타난 바와 같이, 이상 발생 원인 분석 및 출력 단계(S340)에서, 모니터링 모듈(130)은 모니터링 대상 데이터의 이상이 발생한 시점(t1)의 변수비율에서 각 시계열 데이터에 대한 변수 중요도와, 모니터링 대상 데이터의 이상이 발생하지 않은 시점(t2)의 변수비율에서 각 시계열 데이터에 대한 변수 중요도를 비교함으로써 이상이 발생한 시계열 데이터를 판단할 수 있다.

도 5의 예에서는 이상이 발생한 시점과 정상인 시점의 변수 비율을 비교할 때, 여섯번째 변수(v6)의 변수 중요도가 사전 설정된 기준 이상 크게 차이가 난다. 이러한 결과를 기반으로 여섯번째 변수, 즉 총 k 개 파라미터에 대한 시계열 데이터 중 6번 파라미터에 대한 시계열 데이터의 입력에 의해 이상이 발생하였음을 확인할 수 있다.

이상에서 설명한 바와 같이, 본 발명의 여러 실시형태에 따른 복수의 시계열 데이터의 모니터링 장치 및 방법은 복수의 시계열 데이터를 하나의 프로파일 데이터로 전처리하고 기계 학습을 수행하여 얻어진 정보들을 이용하여 모니터링 차트를 구축함으로써, 복수의 파라미터에 대한 데이터를 하나의 지표로 관리하고, 데이터의 이상 여부를 모니터링할 수 있다.

특히 본 발명의 여러 실시형태에 따른 복수의 시계열 데이터의 모니터링 장치 및 방법은, 데이터의 이상이 발생한 경우 각 시점에서 정의된 변수 비율을 이용하여 변수 비율에서 특정 시계열 데이터가 차지하는 변수 중요도를 판단함으로써 이상이 발생한 시점에서 이상의 원인이 되는 변수를 정확하게 판별할 수 있다.

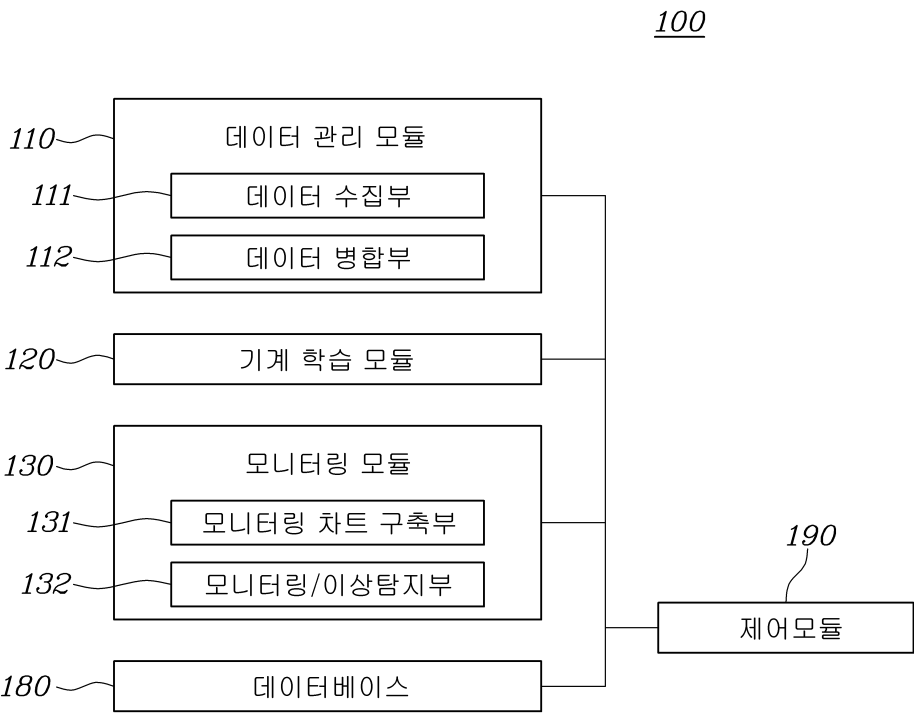
[0113] 이상에서 본 발명의 특정한 실시형태에 관련하여 도시하고 설명하였지만, 이하의 특허청구범위에 의해 제공되는 본 발명의 기술적 사상을 벗어나지 않는 한도 내에서, 본 발명이 다양하게 개량 및 변화될 수 있다는 것은 당 기술분야에서 통상의 지식을 가진 자에게 있어서 자명할 것이다.

부호의 설명

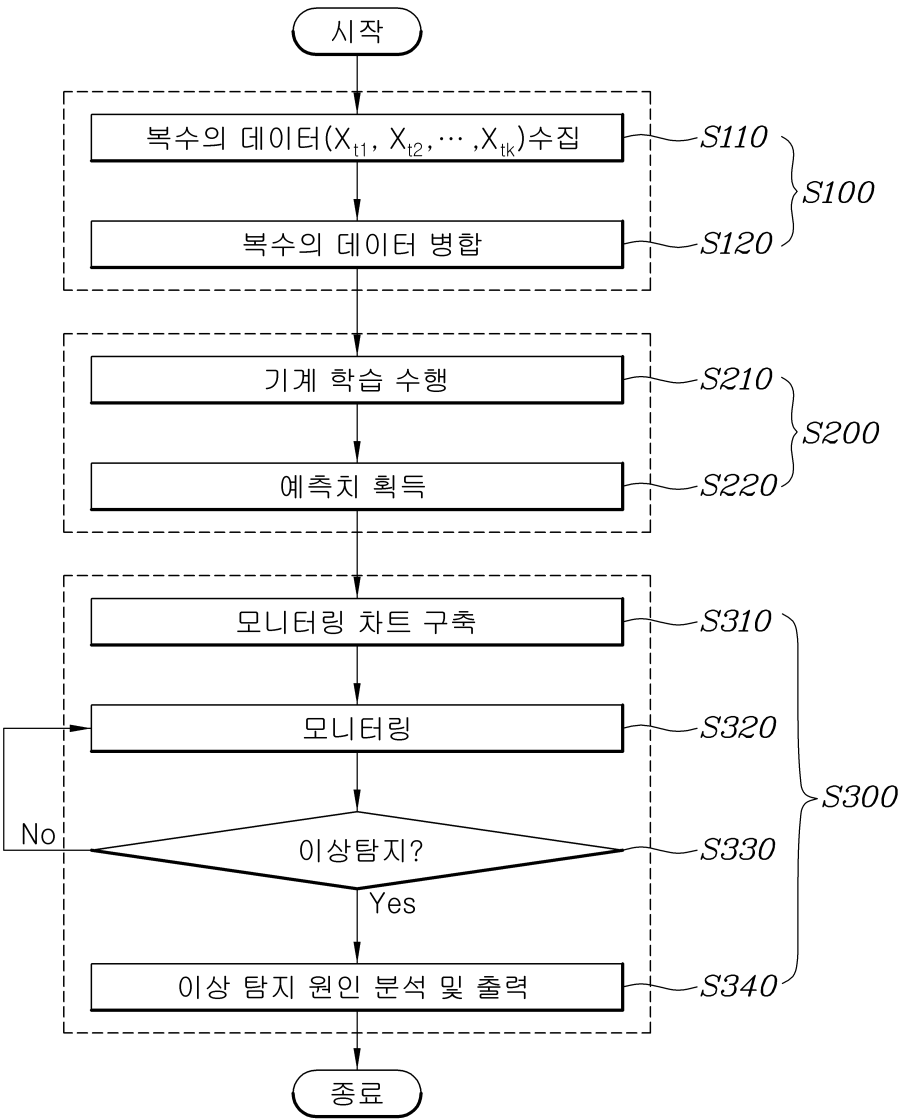
- [0115]
- | | |
|-----------------|------------------|
| 110: 데이터 관리 모듈 | 111: 데이터 수집부 |
| 112: 데이터 병합부 | 120: 기계 학습 모듈 |
| 130: 모니터링 모듈 | 131: 모니터링 차트 구축부 |
| 132: 모니터링/이상탐지부 | 180: 데이터베이스 |
| 190: 제어모듈 | |

도면

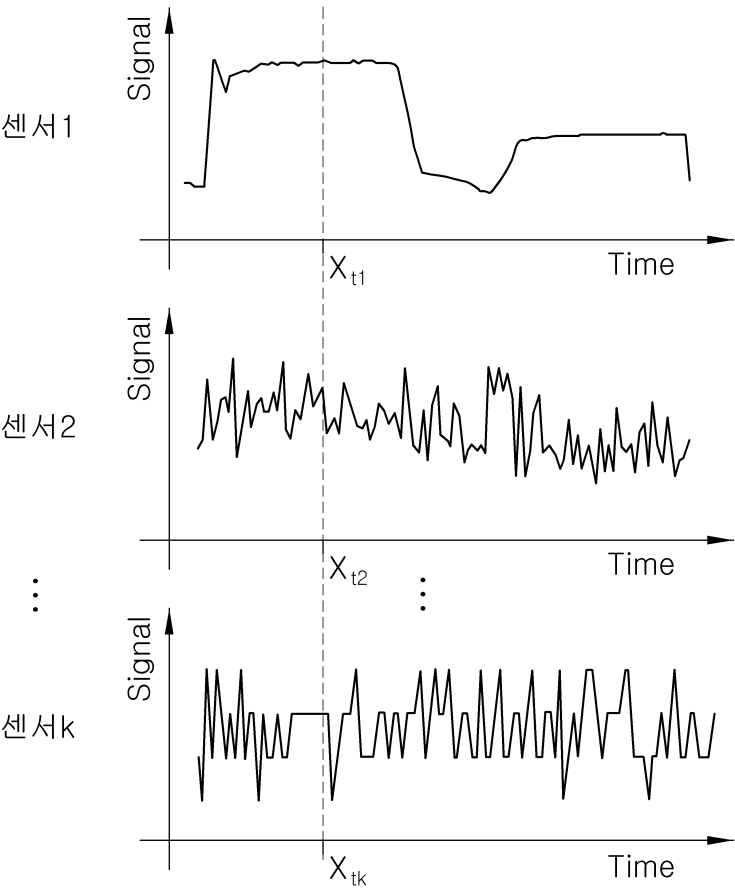
도면1



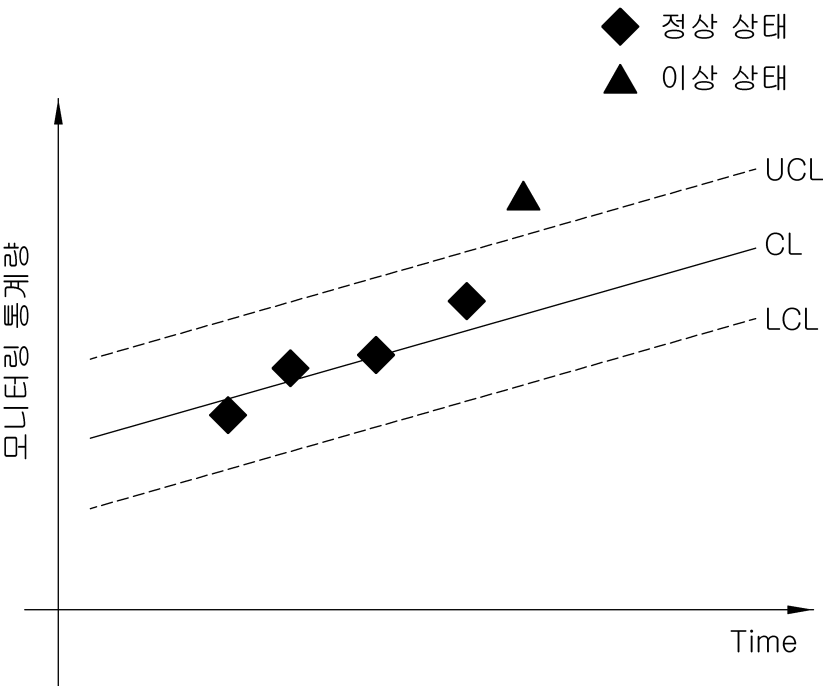
도면2



도면3



도면4



도면5

