(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2025-0031027
(43) 공개일자 2025년03월06일

(51) 국제특허분류(Int. Cl.)
G06F 11/30 (2006.01) *G06F 11/22* (2017.01)
G06F 17/14 (2006.01) *G06N 20/00* (2019.01)
G06N 3/04 (2023.01)
(52) CPC특허분류
G06F 11/3072 (2013.01)
G06F 11/2263 (2013.01)
(21) 출원번호 10-2023-0111086
(22) 출원일자 2023년08월24일
심사청구일자 2023년08월24일

(71) 출원인
고려대학교 산학협력단
서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)
(72) 발명자
이성환
서울특별시 강남구 언주로30길 13, 비동 2902호 (도곡동, 대림아크로빌, 대림아크로텔)
조하영
서울특별시 구로구 디지털로31길 120 (구로동 한신휴플러스아파트) 105동 802호
(74) 대리인
윤귀상

전체 청구항 수 : 총 16 항

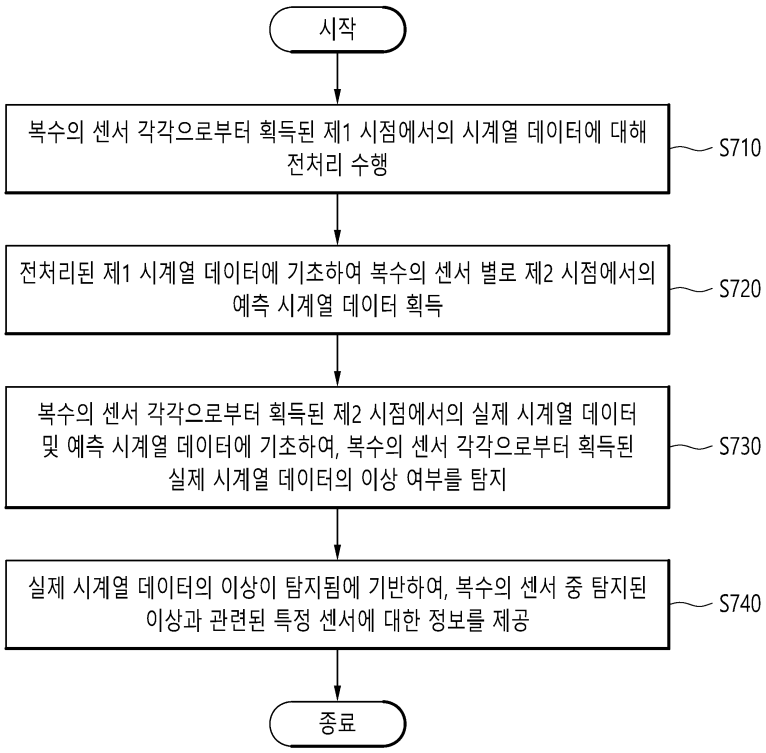
(54) 발명의 명칭 **노드 조건 모듈에 기초한 그래프 네트워크를 이용한 다변량 시계열 데이터 이상 탐지 방법 및 장치**

(57) 요 약

그래프 네트워크를 이용하여 다변량 시계열 데이터 이상을 탐지하는 방법 및 장치가 개시된다. 본 개시의 일 실시예에 따른, 장치에 의해 수행되는, 그래프 네트워크를 이용하여 다변량 시계열 데이터 이상을 탐지하는 방법은, 복수의 센서 각각으로부터 획득된 제1 시점에서의 제1 시계열 데이터에 대해 전처리를 수행하는 단계;

(뒷면에 계속)

대 표 도 - 도7



상기 전처리된 제1 시계열 데이터에 기초하여 상기 복수의 센서 별로 제2 시점에서의 예측 시계열 데이터를 획득하는 단계; 상기 복수의 센서 각각으로부터 획득된 상기 제2 시점에서의 실제 시계열 데이터 및 상기 제2 시점에서의 예측 시계열 데이터에 기초하여, 상기 복수의 센서 각각으로부터 획득된 제2 시점에서의 실제 시계열 데이터의 이상 여부를 탐지하는 단계; 상기 실제 시계열 데이터의 이상이 탐지됨에 기반하여, 상기 복수의 센서 중 상기 탐지된 이상과 관련된 특정 센서에 대한 정보를 제공하는 단계를 포함하고, 상기 특정 센서에 대한 정보는, 상기 복수의 센서 중 상기 특정 센서와 관련된 적어도 하나의 센서 및 상기 특정 센서 간의 관련도를 나타내는 그래프 정보를 포함할 수 있다.

(52) CPC특허분류
G06F 11/3089 (2013.01)
G06F 17/142 (2013.01)
G06N 20/00 (2021.08)
G06N 3/04 (2023.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711193973
과제번호	2019-0-00079-005
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	정보통신방송혁신인재양성
연구과제명	인공지능대학원지원(고려대학교)
기 여 율	1/1
과제수행기관명	고려대학교산학협력단
연구기간	2023.01.01 ~ 2023.12.31

명 세 서

청구범위

청구항 1

장치에 의해 수행되는, 그래프 네트워크를 이용하여 다변량 시계열 데이터 이상을 탐지하는 방법에 있어서, 상기 방법은:

복수의 센서 각각으로부터 획득된 제1 시점에서의 제1 시계열 데이터에 대해 전처리를 수행하는 단계;

상기 전처리된 제1 시계열 데이터에 기초하여 상기 복수의 센서 별로 제2 시점에서의 예측 시계열 데이터를 획득하는 단계;

상기 복수의 센서 각각으로부터 획득된 상기 제2 시점에서의 실제 시계열 데이터 및 상기 제2 시점에서의 예측 시계열 데이터에 기초하여, 상기 복수의 센서 각각으로부터 획득된 상기 제2 시점에서의 실제 시계열 데이터의 이상 여부를 탐지하는 단계;

상기 제2 시점에서의 실제 시계열 데이터의 이상이 탐지됨에 기반하여, 상기 복수의 센서 중 상기 탐지된 이상과 관련된 특정 센서에 대한 정보를 제공하는 단계를 포함하고,

상기 특정 센서에 대한 정보는, 상기 복수의 센서 중 상기 특정 센서와 관련된 적어도 하나의 센서 및 상기 특정 센서 간의 관련도를 나타내는 그래프 정보를 포함하는, 방법.

청구항 2

제1항에 있어서,

상기 제2 시점에서의 예측 시계열 데이터를 획득하는 단계는,

상기 복수의 센서 각각을 복수의 노드로서 설정한 특정 그래프를 생성하고, 상기 전처리된 제1 시계열 데이터를 인코딩하여 상기 복수의 센서 각각에 대응하는 노드의 표현(representation)을 획득하는 단계;

상기 노드의 표현 및 상기 특정 그래프에 기초하여, 상기 복수의 센서 각각에 대응되는 노드의 표현을 업데이트하는 단계; 및

상기 업데이트된 복수의 센서 각각에 대응되는 노드의 표현으로부터 상기 제2 시점에서의 예측 시계열 데이터를 획득하는 단계를 포함하는, 방법.

청구항 3

제2항에 있어서,

상기 복수의 센서 각각에 대응하는 노드의 표현을 획득하는 단계는,

상기 복수의 센서 각각에 대응되는 노드 및 상기 복수의 센서 각각에 대응되는 노드를 대표하는 노드 임베딩 벡터를 획득하는 단계;

상기 노드 임베딩 벡터에 기초하여 상기 복수의 센서 각각에 대응되는 노드 간의 연결을 결정하는 단계; 및

상기 전처리된 제1 시계열 데이터를 인코딩하여 상기 복수의 센서에 대응 되는 노드의 표현을 획득하는 단계를 포함하는, 방법.

청구항 4

제2항에 있어서,

상기 특정 그래프는,

상기 복수의 센서 중 제1 센서에 대응되는 제1 노드 및 상기 제1 노드와 연결된 제2 센서에 대응되는 제2 노드를 포함하고,

상기 복수의 센서 각각에 대응되는 노드의 표현을 업데이트하는 단계는,

상기 제1 노드의 임베딩 벡터 및 상기 제2 노드의 임베딩 벡터에 기초하여 상기 제2 노드의 표현을 변환하여 상기 제1 노드의 표현을 업데이트하는 단계를 포함하는, 방법

청구항 5

제3항에 있어서,

상기 복수의 센서 각각에 대응되는 노드의 표현으로부터 상기 제2 시점에서의 예측 시계열 데이터를 획득하는 단계는,

상기 제1 센서에 대응되는 제1 노드의 임베딩 벡터에 기초하여 상기 업데이트 된 제1 노드의 표현을 디코딩(decoding)함으로써 제1 센서에 대한 상기 제2 시점에서의 예측 시계열 데이터를 획득하는 단계를 포함하는, 방법.

청구항 6

제1항에 있어서,

상기 복수의 센서 각각으로부터 획득된 상기 제2 시점에서의 실제 시계열 데이터의 이상 여부를 탐지하는 단계는,

상기 복수의 센서 각각으로부터 획득된 상기 제2 시점에서의 실제 시계열 데이터 및 상기 제2 시점에서의 예측 시계열 데이터를 이용하여 상기 복수의 센서 별 이상 점수를 산출하는 단계;

상기 복수의 센서 별 이상 점수에 대해 후처리(post processing)를 수행하는 단계; 및

상기 후처리된 복수의 센서 별 이상 점수에 기초하여 상기 복수의 센서 별 최종 이상 점수를 산출하는 단계; 및

상기 복수의 센서 별 최종 이상 점수를 통해 상기 복수의 센서 각각으로부터 획득된 상기 제2 시점에서의 실제 시계열 데이터의 이상 여부를 탐지하는 단계를 포함하는, 방법.

청구항 7

제6항에 있어서,

상기 탐지된 이상과 관련된 특정 센서에 대한 정보를 제공하는 단계는,

상기 적어도 하나의 센서가 상기 특정 센서에 영향을 주는 정도에 기초하여 상기 적어도 하나의 센서 각각의 우선순위를 산출하는 단계;

상기 적어도 하나의 센서 각각의 우선순위에 기초하여 상기 적어도 하나의 센서 및 상기 특정 센서 간의 관련도를 산출하는 단계를 포함하는, 방법.

청구항 8

제1항에 있어서,

상기 전처리하는 단계는,

상기 제1 시점에서의 제1 시계열 데이터에 대해 미리 정의된 전처리 방법을 적용하는 단계를 포함하고,

상기 미리 정의된 전처리 방법은,

상기 제1 시점에서의 제1 시계열 데이터에 대해 정규화(normalization) 또는 패스트 푸리에 변환(fast Fourier transformation)을 포함하는, 방법.

청구항 9

그래프 네트워크를 이용하여 다변량 시계열 데이터 이상을 탐지하는 장치에 있어서, 상기 장치는:

복수의 센서 각각으로부터 획득된 제1 시점에서의 제1 시계열 데이터에 대해 전처리를 수행하는 전처리부;

상기 전처리된 제1 시계열 데이터에 기초하여 상기 복수의 센서 별로 제2 시점에서의 예측 시계열 데이터를 획

득하는 예측부;

상기 복수의 센서 각각으로부터 획득된 상기 제2 시점에서의 실제 시계열 데이터 및 상기 제2 시점에서의 예측 시계열 데이터에 기초하여, 상기 복수의 센서 각각으로부터 획득된 상기 제2 시점에서의 실제 시계열 데이터의 이상 여부를 탐지하는 이상 탐지부; 및

상기 제2 시점에서의 실제 시계열 데이터의 이상이 탐지됨에 기반하여, 상기 복수의 센서 중 상기 탐지된 이상과 관련된 특정 센서에 대한 정보를 제공하는 이상 원인 분석부를 포함하고,

상기 특정 센서에 대한 정보는, 상기 복수의 센서 중 상기 특정 센서와 관련된 적어도 하나의 센서 및 상기 특정 센서 간의 관련도를 나타내는 그래프 정보를 포함하는, 장치.

청구항 10

제9항에 있어서,

상기 예측부는,

상기 복수의 센서 각각을 복수의 노드로서 설정한 특정 그래프를 생성하고, 상기 전처리된 제1 시계열 데이터를 인코딩하여 상기 복수의 센서 각각에 대응하는 노드의 표현(representation)을 획득하고;

상기 노드의 표현 및 상기 특정 그래프에 기초하여, 상기 복수의 센서 각각에 대응되는 노드의 표현을 업데이트하고; 및

상기 업데이트된 복수의 센서 각각에 대응되는 노드의 표현으로부터 상기 제2 시점에서의 예측 시계열 데이터를 획득하는, 장치.

청구항 11

제10항에 있어서,

상기 예측부는,

상기 복수의 센서 각각에 대응되는 노드 및 상기 복수의 센서 각각에 대응되는 노드를 대표하는 노드 임베딩 벡터를 획득하고;

상기 노드 임베딩 벡터에 기초하여 상기 복수의 센서 각각에 대응되는 노드 간의 연결을 결정하고; 및

상기 전처리된 제1 시계열 데이터를 인코딩하여 상기 복수의 센서에 대응되는 노드의 표현을 획득하는, 장치.

청구항 12

제10항에 있어서,

상기 특정 그래프는,

상기 복수의 센서 중 제1 센서에 대응되는 제1 노드 및 상기 제1 노드와 연결된 제2 센서에 대응되는 제2 노드를 포함하고,

상기 예측부는,

상기 제1 노드의 임베딩 벡터 및 상기 제2 노드의 임베딩 벡터에 기초하여 상기 제2 노드의 표현을 변환하여 상기 제1 노드의 표현을 업데이트하는, 장치.

청구항 13

제11항에 있어서,

상기 제1 센서에 대응되는 제1 노드의 임베딩 벡터에 기초하여 상기 업데이트된 제1 노드의 노드 표현을 디코딩(decoding)함으로써 제1 센서에 대한 상기 제2 시점에서의 예측 시계열 데이터가 획득되는, 장치.

청구항 14

제9항에 있어서,

상기 이상 탐지부는,

상기 복수의 센서 각각으로부터 획득된 상기 제2 시점에서의 실제 시계열 데이터 및 상기 제2 시점에서의 예측 시계열 데이터를 이용하여 상기 복수의 센서 별 이상 점수를 산출하고;

상기 복수의 센서 별 이상 점수에 대해 후처리(post processing)를 수행하고; 및

상기 후처리된 복수의 센서 별 이상 점수에 기초하여 상기 복수의 센서 별 최종 이상 점수를 산출하고; 및

상기 복수의 센서 별 최종 이상 점수를 통해 상기 복수의 센서 각각으로부터 획득된 상기 제2 시점에서의 실제 시계열 데이터의 이상 여부를 탐지하는, 장치.

청구항 15

제14항에 있어서,

상기 이상 원인 분석부는,

상기 적어도 하나의 센서가 상기 특정 센서에 영향을 주는 정도에 기초하여 상기 적어도 하나의 센서 각각의 우선순위를 산출하고; 및

상기 적어도 하나의 센서 각각의 우선순위에 기초하여 상기 적어도 하나의 센서 및 상기 특정 센서 간의 관련도를 산출하는, 장치.

청구항 16

제9항에 있어서,

상기 전처리부는,

상기 제1 시점에서의 제1 시계열 데이터에 대해 미리 정의된 전처리 방법을 적용하고,

상기 미리 정의된 전처리 방법은,

상기 제1 시점에서의 제1 시계열 데이터에 대해 정규화(normalization) 또는 패스트 푸리에 변환(fast Fourier transformation)을 포함하는, 장치.

발명의 설명

기 술 분 야

[0001] 본 개시는 다변량 시계열 데이터의 이상 탐지 기술에 관한 것이며, 보다 상세하게는 센서 사이의 연결 관계를 모르는 시스템 상에서 획득된 다변량 시계열 데이터의 이상을 탐지하는 방법 및 장치에 관한 것이다.

배 경 기 술

[0002] IoT(internet of things)의 발달을 통해서 복수의 유형의 센서 데이터를 다양한 분야에서 획득할 수 있게 되었다. IoT 기반 시스템에서 획득된 센서 데이터는 모니터링 및 분석을 통해서 시스템의 정상 작동 여부를 판단하거나 외부의 공격 여부를 판단하기 위하여 사용될 수 있다. 다만, 많은 유형의 센서들로부터 짧은 시간에 방대한 데이터가 생성되는 바, 사용자가 실시간으로 이를 직접 모니터링하고 분석하는 것은 쉽지 않은 일이다.

[0003] 한편, 인공 지능(artificial intelligence, AI)은 많은 데이터를 통하여 특정 데이터베이스에 함축/내포된 정보를 자동으로 추출하고 가공하는데 특화가 되어 있다. 상술된 AI의 특성에 기초하여 시계열 데이터를 이용하여 모니터링 및 분석을 위한 이상 탐지를 하는 여러 방법들이 고안되어 왔다.

[0004] 기존에는 주로 오토 인코더 기반의 방법이 적용됨에 따라, 입력과 출력을 최대한 같게 학습함으로써 정상 학습 데이터의 특징을 학습하고, 얼마나 잘 재구축 되는지를 통하여 이상 여부를 판별하는 방식이 고안 및 활용되었다.

[0005] 다만, 실제 세계에서는 이상 탐지에는 이상의 발생 유무를 탐지하는 것 뿐만 아니라 탐지된 이상의 원인이 되는지 아는 것이 중요하다. 이상 원인을 빠르게 파악하게 될 경우, 빠른 사후 조치를 통하여 시스템을 원상복구 하는데 걸리는 시간을 줄일 수 있다는 장점이 존재한다.

[0006] 기존의 기술의 경우 단순히 데이터의 이상 발생 여부만을 판단하기 때문에, 센서의 고장이 정상 작동 범위를 벗어나게 하는 작동을 하여 해당 센서에서 이상을 발생시키는 경우, 이상의 원인이 되는 센서를 즉각적으로 찾아서 사후처리를 할 수 있다.

[0007] 다만, 센서의 고장이 정상 작동 범위 내에서 작동하여 다른 센서에서 이상을 발생시키는 경우 이상이 발생된 센서와 원인이 되는 센서가 다르기 때문에 이를 발견하기 힘들다.

[0008] 상술된 기존 기술의 문제점을 해결하기 위하여 센서 간의 연결 구조를 이용한 그래프를 이용하여 이상을 탐지하고 그래프 구조를 이용하여 원인이 될 수 있는 센서들을 알려주는 모델들이 등장하였다.

[0009] 다만, 상술된 그래프 기반의 이상 탐지 기술은 센서 간의 연결 구조를 알아야 적용될 수 있다는 한계점이 존재하였다. 간단한 시스템의 경우 연결 구조를 파악하는 것이 어렵지 않을 수 있다. 다만, 대규모 발전소, 수자원 분배 시스템과 같이 대규모 기간 산업에 사용되는 시스템의 경우 규모가 크며, 여러 시스템들이 복합적으로 얽혀있는 바, 이런 복잡한 시스템에서 측정되는 센서들의 관계를 알아낸다는 것은 거의 불가능하다는 한계가 존재한다.

발명의 내용

해결하려는 과제

[0010] 본 개시는 상술된 문제점을 해결하기 위해 안출된 것으로서, 본 개시의 목적은 센서 간의 연결 구조를 모르는 다변량 시계열 데이터에서 학습된 센서의 임베딩 벡터를 이용하여 연결 구조를 식별하면서 시계열의 이상을 탐지하는 방법 및 장치를 제공함에 있다.

[0011] 또한, 본 개시의 목적은 센서의 임베딩 벡터를 이용하여 센서 간의 연결 구조를 학습하고 미래의 값을 예측하는 그래프 기반 신경망을 구축하고, 이를 통하여 다변량 데이터의 이상을 탐지하고, 이상을 유발하는 센서의 중요도를 제공하는 방법 및 장치를 제공함에 있다.

[0012] 본 개시의 목적들은 이상에서 언급한 목적으로 제한되지 않으며, 언급되지 않은 본 발명의 다른 목적 및 장점들은 하기의 설명에 의해서 이해될 수 있고, 본 개시의 실시예에 의해 보다 분명하게 이해될 것이다. 또한, 본 개시의 목적 및 장점들은 청구범위에 나타낸 수단 및 그 조합에 의해 실현될 수 있음을 쉽게 알 수 있을 것이다.

과제의 해결 수단

[0013] 본 개시의 다양한 실시예로, 장치에 의해 수행되는, 그래프 네트워크를 이용하여 다변량 시계열 데이터 이상을 탐지하는 방법은, 복수의 센서 각각으로부터 획득된 제1 시점에서의 제1 시계열 데이터에 대해 전처리를 수행하는 단계; 상기 전처리된 제1 시계열 데이터에 기초하여 상기 복수의 센서 별로 제2 시점에서의 예측 시계열 데이터를 획득하는 단계; 상기 복수의 센서 각각으로부터 획득된 상기 제2 시점에서의 실제 시계열 데이터 및 상기 제2 시점에서의 예측 시계열 데이터에 기초하여, 상기 복수의 센서 각각으로부터 획득된 상기 제2 시점에서의 실제 시계열 데이터의 이상 여부를 탐지하는 단계; 상기 제2 시점에서의 실제 시계열 데이터의 이상이 탐지됨에 기반하여, 상기 복수의 센서 중 상기 탐지된 이상과 관련된 특정 센서에 대한 정보를 제공하는 단계를 포함하고, 상기 특정 센서에 대한 정보는, 상기 복수의 센서 중 상기 특정 센서와 관련된 적어도 하나의 센서 및 상기 특정 센서 간의 관련도를 나타내는 그래프 정보를 포함할 수 있다.

[0014] 그리고, 상기 제2 시점에서의 예측 시계열 데이터를 획득하는 단계는, 상기 복수의 센서 각각을 복수의 노드로서 설정한 특정 그래프를 생성하고, 상기 전처리된 제1 시계열 데이터를 인코딩하여 상기 복수의 센서 각각에 대응하는 노드의 표현(representation)을 획득하는 단계; 상기 노드의 표현 및 상기 특정 그래프에 기초하여, 상기 복수의 센서 각각에 대응되는 노드의 표현을 업데이트하는 단계; 및 상기 업데이트된 복수의 센서 각각에 대응되는 노드의 표현으로부터 상기 예측 시계열 데이터를 획득하는 단계를 포함할 수 있다.

[0015] 그리고, 상기 복수의 센서 각각에 대응하는 노드의 표현을 획득하는 단계는, 상기 복수의 센서 각각에 대응되는 노드 및 상기 복수의 센서 각각에 대응되는 노드를 대표하는 노드 임베딩 벡터를 획득하는 단계; 상기 노드 임베딩 벡터에 기초하여 상기 복수의 센서 각각에 대응되는 노드 간의 연결을 결정하는 단계; 및 상기 전처리된 제1 시계열 데이터를 인코딩하여 상기 복수의 센서에 대응 되는 노드의 표현을 획득하는 단계를 포함할 수 있다.

[0016] 그리고, 상기 특정 그래프는, 상기 복수의 센서 중 제1 센서에 대응되는 제1 노드 및 상기 제1 노드와 연결된

제2 센서에 대응되는 제2 노드를 포함하고, 상기 복수의 센서 각각에 대응되는 노드의 표현을 업데이트하는 단계는, 상기 제1 노드의 임베딩 벡터 및 상기 제2 노드의 임베딩 벡터에 기초하여 상기 제2 노드의 표현을 변환하여 상기 제1 노드의 표현을 업데이트하는 단계를 포함할 수 있다.

[0017] 그리고, 상기 복수의 센서 각각에 대응되는 노드의 표현으로부터 상기 예측 시계열 데이터를 획득하는 단계는, 상기 제1 센서에 대응되는 제1 노드의 임베딩 벡터에 기초하여 상기 업데이트 된 제1 노드의 노드 표현을 디코딩(decoding)함으로써 제1 센서에 대한 상기 제2 시점에서의 예측 시계열 데이터가 획득될 수 있다.

[0018] 그리고, 상기 복수의 센서 각각으로부터 획득된 상기 제2 시점에서의 실제 시계열 데이터의 이상 여부를 탐지하는 단계는, 상기 복수의 센서 각각으로부터 획득된 상기 제2 시점에서의 실제 시계열 데이터 및 상기 제2 시점에서의 예측 시계열 데이터를 이용하여 상기 복수의 센서 별 이상 점수를 산출하는 단계; 상기 복수의 센서 별 이상 점수에 대해 후처리(post processing)를 수행하는 단계; 및 상기 후처리된 복수의 센서 별 이상 점수에 기초하여 상기 복수의 센서 별 최종 이상 점수를 산출하는 단계; 및 상기 복수의 센서 별 최종 이상 점수를 통해 상기 복수의 센서 각각으로부터 획득된 상기 제2 시점에서의 실제 시계열 데이터의 이상 여부를 탐지하는 단계를 포함할 수 있다.

[0019] 그리고, 상기 탐지된 이상과 관련된 특정 센서에 대한 정보를 제공하는 단계는, 상기 적어도 하나의 센서가 상기 특정 센서에 영향을 주는 정도에 기초하여 상기 적어도 하나의 센서 각각의 우선순위를 산출하는 단계; 상기 적어도 하나의 센서 각각의 우선순위에 기초하여 상기 적어도 하나의 센서 및 상기 특정 센서 간의 관련도를 산출하는 단계를 포함할 수 있다.

[0020] 그리고, 상기 전처리하는 단계는, 상기 제1 시점에서의 제1 시계열 데이터에 대해 미리 정의된 전처리 방법을 적용하는 단계를 포함하고, 상기 미리 정의된 전처리 방법은, 상기 제1 시점에서의 제1 시계열 데이터에 대해 정규화(normalization) 또는 패스트 푸리에 변환(fast Fourier transformation) 수행하는 단계를 포함할 수 있다.

[0021] 본 개시의 또 다른 예로, 그래프 네트워크를 이용하여 다변량 시계열 데이터 이상을 탐지하는 장치는, 복수의 센서 각각으로부터 획득된 제1 시점에서의 제1 시계열 데이터에 대해 전처리를 수행하는 전처리부; 상기 전처리된 제1 시계열 데이터에 기초하여 상기 복수의 센서 별로 제2 시점에서의 예측 시계열 데이터를 획득하는 예측부; 상기 복수의 센서 각각으로부터 획득된 상기 제2 시점에서의 실제 시계열 데이터 및 상기 제2 시점에서의 예측 시계열 데이터에 기초하여, 상기 복수의 센서 각각으로부터 획득된 상기 제2 시점에서의 실제 시계열 데이터의 이상 여부를 탐지하는 이상 탐지부; 및 상기 제2 시점에서의 실제 시계열 데이터의 이상이 탐지됨에 기반하여, 상기 복수의 센서 중 상기 탐지된 이상과 관련된 특정 센서에 대한 정보를 제공하는 이상 원인 분석부를 포함하고, 상기 특정 센서에 대한 정보는, 상기 복수의 센서 중 상기 특정 센서와 관련된 적어도 하나의 센서 및 상기 특정 센서 간의 관련도를 나타내는 그래프 정보를 포함할 수 있다.

[0022] 그리고, 상기 예측부는, 상기 복수의 센서 각각을 복수의 노드로서 설정한 특정 그래프를 생성하고, 상기 전처리된 제1 시계열 데이터를 인코딩하여 상기 복수의 센서 각각에 대응하는 노드의 표현(representation)을 획득하고; 상기 노드의 표현 및 상기 특정 그래프에 기초하여, 상기 복수의 센서 각각에 대응되는 노드의 표현을 업데이트하고; 및 상기 업데이트된 복수의 센서 각각에 대응되는 노드의 표현으로부터 상기 제2 시점에서의 예측 시계열 데이터를 획득할 수 있다.

[0023] 상기 예측부는, 상기 복수의 센서 각각에 대응되는 노드 및 상기 복수의 센서 각각에 대응되는 노드를 대표하는 노드 임베딩 벡터를 획득하고; 상기 노드 임베딩 벡터에 기초하여 상기 복수의 센서 각각에 대응되는 노드 간의 연결을 결정하고; 및 상기 전처리된 제1 시계열 데이터를 인코딩하여 상기 복수의 센서에 대응되는 노드의 표현을 획득할 수 있다.

[0024] 그리고, 상기 예측부는, 상기 제1 노드의 임베딩 벡터 및 상기 제2 노드의 임베딩 벡터에 기초하여 상기 제2 노드의 표현을 변환하여 상기 제1 노드의 표현을 업데이트할 수 있다.

[0025] 그리고, 상기 이상 탐지부는, 상기 복수의 센서 각각으로부터 획득된 상기 제2 시점에서의 실제 시계열 데이터 및 상기 제2 시점에서의 예측 시계열 데이터를 이용하여 상기 복수의 센서 별 이상 점수를 산출하고; 상기 복수의 센서 별 이상 점수에 대해 후처리(post processing)를 수행하고; 및 상기 후처리된 복수의 센서 별 이상 점수에 기초하여 상기 복수의 센서 별 최종 이상 점수를 산출하고; 및 상기 복수의 센서 별 최종 이상 점수를 통해 상기 복수의 센서 각각으로부터 획득된 제2 시점에서의 실제 시계열 데이터의 이상 여부를 탐지할 수 있다.

[0026] 그리고, 상기 이상 원인 분석부는, 상기 적어도 하나의 센서가 상기 특정 센서에 영향을 주는 정도에 기초하여

상기 적어도 하나의 센서 각각의 우선순위를 산출하고; 및 상기 적어도 하나의 센서 각각의 우선순위에 기초하여 상기 적어도 하나의 센서 및 상기 특정 센서 간의 관련도를 산출할 수 있다.

[0027] 그리고, 상기 전처리부는, 상기 제1 시점에서의 제1 시계열 데이터에 대해 미리 정의된 전처리 방법을 적용하고, 상기 미리 정의된 전처리 방법은, 상기 제1 시점에서의 제1 시계열 데이터에 대해 정규화(normalization) 또는 패스트 푸리에 변환(fast Fourier transformation)을 포함할 수 있다.

[0028] 상기 과제의 해결 수단은, 본 개시의 특징을 모두 열거한 것은 아니다. 본 개시의 다양한 특징과 그에 따른 장점과 효과는 아래의 구체적인 실시예를 참조하여 보다 상세하게 이해될 수 있을 것이다.

발명의 효과

[0029] 본 개시의 다양한 실시예에 의해, 여러 센서를 통해 출력된 다변량 시계열 데이터를 통하여 센서들 간의 관계를 찾아내고, 이를 기반으로 다변량 시계열 이상 탐지를 수행하며, 이상 발생의 원인이 되는 센서를 특정하는 방법 및 장치가 개시될 수 있다.

[0030] 또한, 본 개시의 다양한 실시예에 의해, 데이터의 이상 탐지 뿐만 아니라 이상의 원인이 되는 센서에 대한 정보를 제공함으로써, 이상 탐지 후 원인 파악에 걸리는 시간이 줄어들 수 있으며 보다 빠른 시스템의 복구가 가능할 수 있다.

[0031] 본 개시의 효과들은 이상에서 언급된 효과로 제한되지 않으며, 언급되지 않은 또 다른 효과들은 아래의 기재로부터 통상의 기술자에게 명확하게 이해될 수 있을 것이다.

도면의 간단한 설명

[0032] 도 1은 다변량 시계열 데이터 이상 탐지 방법을 수행하는 장치의 구성을 도시한 블록도이다.

도 2는 본 개시의 일 실시예에 따른, 전처리부가 수행하는 동작을 설명하기 위한 순서도이다.

도 3은 본 개시의 일 실시예에 따른, 예측부가 수행하는 동작을 설명하기 위한 순서도이다.

도 4는 본 개시의 일 실시예로, 인코딩된 벡터 및 그래프 구조를 이용하여 간선 별 노드 업데이트하는 과정을 설명하기 위한 도면이다.

도 5는 본 개시의 일 실시예에 따른, 이상 탐지부가 수행하는 동작을 설명하기 위한 도면이다.

도 6은 본 개시의 일 실시예에 따른, 이상 원인 분석부가 수행하는 동작을 설명하기 위한 도면이다.

도 7은 본 개시의 일 실시예에 따른, 그래프 네트워크를 이용하여 다변량 시계열 데이터 이상을 탐지하는 방법을 설명하기 위한 순서도이다.

도 8은 본 개시의 일 실시예에 따른, 인코딩된 입력과 그래프 구조를 이용하여 노드의 표현을 업데이트하는 방법을 설명하기 위한 도면이다.

발명을 실시하기 위한 구체적인 내용

[0033] 본 개시의 이점 및 특징, 그리고 그것들을 달성하는 방법은 첨부되는 도면과 함께 상세하게 후술되어 있는 실시예들을 참조하면 명확해질 것이다. 그러나, 본 개시는 이하에서 개시되는 실시예들에 제한되는 것이 아니라 서로 다른 다양한 형태로 구현될 수 있으며, 단지 본 실시예들은 본 개시의 개시가 완전하도록 하고, 본 개시가 속하는 기술 분야의 통상의 기술자에게 본 개시의 범주를 완전하게 알려주기 위해 제공되는 것이며, 본 개시는 청구항의 범주에 의해 정의될 뿐이다.

[0034] 본 명세서에서 사용된 용어는 실시예들을 설명하기 위한 것이며 본 개시를 제한하고자 하는 것은 아니다. 본 명세서에서, 단수형은 문구에서 특별히 언급하지 않는 한 복수형도 포함한다. 명세서에서 사용되는 "포함한다(comprises)" 및/또는 "포함하는(comprising)"은 언급된 구성요소 외에 하나 이상의 다른 구성요소의 존재 또는 추가를 배제하지 않는다.

[0035] 명세서 전체에 걸쳐 동일한 도면 부호는 동일한 구성 요소를 지칭하며, "및/또는"은 언급된 구성요소들의 각각 및 하나 이상의 모든 조합을 포함한다. 비록 "제1", "제2" 등이 다양한 구성요소들을 서술하기 위해서 사용되나, 이들 구성요소들은 이들 용어에 의해 제한되지 않음은 물론이다. 이들 용어들은 단지 하나의 구성요소를 다른 구성요소와 구별하기 위하여 사용하는 것이다. 따라서, 이하에서 언급되는 제1 구성요소는 본 개시의

기술적 사상 내에서 제2 구성요소일 수도 있음은 물론이다.

[0036] 다른 정의가 없다면, 본 명세서에서 사용되는 모든 용어(기술 및 과학적 용어를 포함)는 본 개시가 속하는 기술 분야의 통상의 기술자에게 공통적으로 이해될 수 있는 의미로 사용될 수 있을 것이다. 또한, 일반적으로 사용되는 사전에 정의되어 있는 용어들은 명백하게 특별히 정의되어 있지 않는 한 이상적으로 또는 과도하게 해석되지 않는다.

[0037] 공간적으로 상대적인 용어인 "아래(below)", "아래(beneath)", "하부(lower)", "위(above)", "상부(upper)" 등은 도면에 도시되어 있는 바와 같이 하나의 구성요소와 다른 구성요소과의 상관관계를 용이하게 기술하기 위해 사용될 수 있다. 공간적으로 상대적인 용어는 도면에 도시되어 있는 방향에 더하여 사용시 또는 동작 시 구성요소들의 서로 다른 방향을 포함하는 용어로 이해되어야 한다.

[0038] 예를 들어, 도면에 도시되어 있는 구성요소를 뒤집을 경우, 다른 구성요소의 "아래(below)"또는 "아래(beneath)"로 기술된 구성요소는 다른 구성요소의 "위(above)"에 놓여질 수 있다. 따라서, 예시적인 용어인 "아래"는 아래와 위의 방향을 모두 포함할 수 있다. 구성요소는 다른 방향으로도 배향될 수 있으며, 이에 따라 공간적으로 상대적인 용어들은 배향에 따라 해석될 수 있다.

[0039] 이하에서는 도면을 참조하여 노드 조건 모듈에 기초한 그래프 네트워크를 이용한 다변량 시계열 데이터 이상 탐지 방법 및 장치에 대해 구체적으로 설명하도록 한다.

[0041] 본 개시의 일 실시예로, 도 1은 다변량 시계열 데이터 이상 탐지 방법을 수행하는 장치의 구성을 도시한 블록도이다. 도 1에서 설명하는 장치는 스마트 폰, 태블릿 PC, 웨어러블 장치, 노트북 등을 포함할 수 있으나 이에 제한되는 것은 아니다.

[0042] 도 1에 도시된 바와 같이, 장치(100)는 메모리(110), 통신 모듈(120), 디스플레이(130), 및 프로세서(140)를 포함할 수 있다. 다만, 이에 국한되는 것은 아니며, 장치(100)는 필요한 동작에 따라 당업자 관점에서 자명한 범위 내에서 소프트웨어 및 하드웨어 구성이 수정/추가/생략될 수 있다.

[0043] 메모리(110)는 본 장치(100)의 다양한 기능을 지원하는 데이터와, 제어부의 동작을 위한 프로그램을 저장할 수 있고, 입/출력되는 데이터들을 저장할 있고, 본 장치에서 구동되는 다수의 응용 프로그램(application program 또는 애플리케이션(application)), 본 장치의 동작을 위한 데이터들, 명령어들을 저장할 수 있다. 이러한 응용 프로그램 중 적어도 일부는, 무선 통신을 통해 외부 장치로부터 다운로드 될 수 있다.

[0044] 이러한, 메모리(110)는 플래시 메모리 타입(flash memory type), 하드디스크 타입(hard disk type), SSD 타입(Solid State Disk type), SDD 타입(Silicon Disk Drive type), 멀티미디어 카드 마이크로 타입(multimedia card micro type), 카드 타입의 메모리(예를 들어 SD 또는 XD 메모리 등), 램(random access memory; RAM), SRAM(static random access memory), 롬(read-only memory; ROM), EEPROM(electrically erasable programmable read-only memory), PROM(programmable read-only memory), 자기 메모리, 자기 디스크 및 광디스크 중 적어도 하나의 타입의 저장매체를 포함할 수 있다. 또한, 메모리(110)는 본 장치와는 분리되어 있으나, 유선 또는 무선으로 연결된 데이터베이스가 될 수도 있다.

[0045] 통신 모듈(120)은 외부 장치(예로, 복수의 노드에 대응되는 복수의 센서)와 통신을 가능하게 하는 하나 이상의 구성 요소를 포함할 수 있다. 예를 들어, 통신 모듈(120)은 무선통신 모듈, 유선통신 모듈 또는 위치정보 모듈 중 적어도 하나를 포함할 수 있다.

[0046] 여기서, 무선 통신 모듈은 와이파이(Wifi) 모듈, 와이브로(Wireless broadband) 모듈 외에도, GSM(global System for Mobile Communication), CDMA(Code Division Multiple Access), WCDMA(Wideband Code Division Multiple Access), UMTS(universal mobile telecommunications system), TDMA(Time Division Multiple Access), LTE(Long Term Evolution), 4G, 5G, 6G 등 다양한 무선 통신 방식을 지원하는 무선 통신 모듈을 포함할 수 있다.

[0047] 디스플레이(130)는 본 장치(100)에서 처리되는 정보(예를 들어, 복수의 센서로부터 수신된 다변량 시계열 데이터 등)를 표시(출력)한다. 예를 들어, 디스플레이는 본 장치(100)에서 구동되는 응용 프로그램(일 예로, 어플리케이션)의 실행화면 정보, 또는 이러한 실행화면 정보에 따른 UI(User Interface), GUI(Graphic User Interface) 정보를 표시할 수 있다.

[0048] 프로세서(140)는 본 장치(100) 내의 구성요소들의 동작을 제어하기 위한 알고리즘 또는 알고리즘을 재현한 프로그램에 대한 데이터를 저장하는 메모리, 및 메모리에 저장된 데이터를 이용하여 전술한 동작을 수행하는 적어도 하나의 프로세서(미도시)로 구현될 수 있다. 이때, 메모리와 프로세서는 각각 별개의 칩으로 구현될 수 있다. 또는, 메모리와 프로세서는 단일 칩으로 구현될 수도 있다.

[0049] 즉, 프로세서(140)는 본 장치(100) 내의 구성 요소들을 제어함으로써 장치의 전반적인 동작 및 기능을 제어할 수 있다. 프로세서는 이하의 도면에서 설명되는 본 개시에 따른 다양한 실시 예들을 본 장치 상에서 구현하기 위하여, 위에서 살펴본 구성요소들을 중 어느 하나 또는 복수를 조합하여 제어할 수 있다.

[0050] 본 개시의 일 예로, 도 1에 도시될 바와 같이, 프로세서(140)는 전처리부(140-1), 예측부(140-2), 이상 탐지부(140-3), 및 이상 원인 분석부(140-4)를 포함할 수 있다.

[0051] 전처리부(140-1)은 시스템을 구성하는 복수의 센서들로부터 수신된 측정 값(즉, 다변량 시계열 데이터)을 가공할 수 있다. 즉, 전처리부(140-1)는 측정 값을 모델에 입력하기 위한 데이터 가공 절차를 수행할 수 있다.

[0052] 예측부(140-2)는 입력된 측정 값(즉, 전처리 절차가 적용된 측정 값)으로부터 미래의 측정 값을 예측할 수 있다. 즉, 예측부(140-2)는 제2 시점에서의 예측 시계열 데이터를 획득할 수 있다.

[0053] 이상 탐지부(140-3)는 예측부(140-2)에 의해 제2 시점에서의 예측 시계열 데이터 및 제2 시점에서의 실제 시계열 데이터를 이용하여 제2 시점에서의 실제 시계열 데이터의 이상 여부를 탐지할 수 있다. 즉, 이상 탐지부(140-3)는 제2 시점에서의 예측된 시계열 데이터와 제2 시점에서의 실제 측정된 시계열 데이터 간의 비교 동작을 통해 제2 시점에서의 실제 측정된 시계열 데이터로부터 이상이 탐지되는지 식별할 수 있다.

[0054] 이상 원인 분석부(140-4)는 이상 탐지부(140-3)에 의해 제2 시점에서의 실제 시계열 데이터의 이상이 탐지되었을 때 해당 이상의 원인에 대한 정보를 제공할 수 있다. 즉, 이상 원인 분석부(140-4)는 복수의 센서 중 이상이 탐지된 센서를 분석하기 위한 정보를 제공할 수 있다.

[0055] 각 전처리부(140-1), 예측부(140-2), 이상 탐지부(140-3), 및 이상 원인 분석부(140-4)의 동작은 도 2 내지 도 7을 참조하여 구체적으로 설명하도록 한다.

[0057] 도 2는 본 개시의 일 실시예에 따른, 전처리부가 수행하는 동작을 설명하기 위한 순서도이다.

[0058] 시스템에 포함된 N(N은 1 이상의 자연수) 개의 센서로부터 t 시점에 생성되는 N 차원의 벡터 $x^t \in \mathbb{R}^N$를 예측하기 위하여, 전처리부는 미리 저장된 데이터로부터 정해진 시간 창(time window) 길이 w를 가지는 과거 데이터(즉, 제1 시점의 시계열 데이터로, t 시점보다 시간 창 길이만큼의 과거 데이터) $x \in \mathbb{R}^{w \times N}$ 및 t 시점에서의 측정된 실제 시계열 데이터(즉, 제2 시점의 시계열 데이터로, 모델이 t 시점의 값을 예측한 시계열 데이터와의 비교를 위한 데이터)를 포함하는 입력 데이터를 구성하고, 입력 데이터에 대해 전처리를 수행할 수 있다.

[0059] 이하에서는 본 개시의 설명의 편의를 위하여, t 시점보다 전의 시간 창 길이만큼의 데이터(또는, 과거 데이터)를 제1 시점에서의 제1 시계열 데이터로, t 시점의 실제 측정 데이터(또는, 현재 데이터)를 제2 시점에서의 실제 시계열 데이터로 칭하도록 한다.

[0060] 구체적으로, 전처리부는 N 개의 센서로부터 제1 시점에서의 제1 시계열 데이터를 획득 및 저장할 수 있다(S210).

[0061] 전처리부는 제1 시점에서의 제1 시계열 데이터(즉, 예측을 위한 과거 데이터) 및 제2 시점에서의 실제 시계열 데이터를 포함하고 있는 입력 데이터를 구성할 수 있다(S220).

[0062] 구체적으로, 전처리부는 저장된 시계열 데이터를 이용하여 시간 창을 구성함으로써 제1 시점에서의 제1 시계열 데이터를 생성할 수 있다. 전처리부는 획득된 제2 시점에서의 시계열 데이터를 이용하여 제2 시점에서의 실제 시계열 데이터를 생성할 수 있다. 전처리부는 생성된 제1 시점에서의 제1 시계열 데이터 및 제2 시점에서의 실제 시계열 데이터를 이용하여 입력 데이터를 구성할 수 있다.

[0063] 전처리부는 구성된 입력 데이터에 포함되어 있는, 제1 시점에서의 제1 시계열 데이터 및 제2 시점에서의 실제 시계열 데이터에 대해 전처리를 수행할 수 있다(S230).

[0065] 도 3은 본 개시의 일 실시예에 따른, 예측부가 수행하는 동작을 설명하기 위한 순서도이다.

[0066] 예측부는 제1 시점에서의 제1 시계열 데이터(즉, 예측하려는 시점보다 시간 창 길이 이전까지의 데이터(즉, $x^{(t)} = \{x^{t-1}, x^{t-2}, \cdots, x^{t-w}\}$))를 입력 데이터로 이용하여 x^t를 예측할 수 있다.

[0067] 예측부는 입력된 데이터(또는, 전처리된 데이터)에 대해 인코딩을 수행하고 그래프 구조를 획득할 수 있다(S310).

[0068] 구체적으로, 예측부는 입력된 데이터(예로, $x^{(t)} = \{x^{t-1}, x^{t-2}, \cdots, x^{t-w}\}$)에 대해 인코더($f_{enc}$)를 적용할 수 있다. 즉, 예측부는 입력된 복수의 센서 데이터에 대해 정해진 차원의 인코딩 벡터($z_i^{(t)} = f_{enc}(x_i^{(t)})$)로 인코딩을 수행하고 이를 각 센서에 대응되는 노드의 노드 표현으로 사용할 수 있다.

[0069] 예측부는 학습된 노드 임베딩 벡터(v_i)들의 유사도를 이용하여 가장 높은 유사도를 가지는 k개의 노드를 선택하고, 선택된 k 개의 노드를 이용하여 그래프 구조를 획득할 수 있다.

[0070] 예측부는 그래프를 구성하는 노드 간의 유사도를 획득하기 위하여 수학식 1을 이용할 수 있으나 이에 제한되는 것은 아니다.

수학식 1

[0071] $$\text{Similarity}(i,j) = \frac{v_i \cdot v_j}{|v_i||v_j|}$$

[0072] 예측부는 간선 별로 노드 업데이트를 수행할 수 있다(S320).

[0073] 구체적으로, 예측부는 도 8에 도시된 바와 같이 인코딩된 벡터 및 그래프 구조를 이용하여 간선 별 노드 업데이트를 수행할 수 있다. 즉, 예측부는 연결되어 있는 간선 별로 다른 노드 표현(Node representation)을 이용해서 다음 단계의 예측을 위하여 인코딩 벡터를 업데이트할 수 있다.

[0074] 이를 위하여, i 번째 노드에 연결되어 있는 노드들의 인덱스를 $N(i)$라고 가정한다. 예측부는 i 번째 노드의 노드 표현을 업데이트 하기 위하기 위하여 $N(i)$에 속한 i번째 노드들의 노드 표현을 노드 임베딩 벡터 v_i, v_j들을 조건으로 하는 간선 별 노드 업데이트 모듈 f_{update}을 이용하여, i 번째 노드의 노드 표현을 변환 하여 i 번째 노드의 노드 표현을 업데이트하는 간선 별 노드 업데이트를 수행할 수 있다. 즉, 예측부는 서로 연결되어 있는 노드들의 노드 표현을 노드마다 다르게 변경함으로써 간선 별 업데이트 동작을 수행할 수 있다. 예측부는 수학식 2에 따라 간선 별 노드 업데이트 모듈 f_{update}을 이용하여 i 번째 노드에 대하여 간선 별 노드 업데이트 동작을 수행할 수 있다.

수학식 2

[0075] $$z_i^t = \sigma\left(\sum_{i \in N(i) \cup \{i\}} f_{update}\left(z_j^{(t)}; v_i, v_j\right)\right)$$

[0076] 예측부는 노드 별 디코딩 동작을 수행할 수 있다(S330).

[0077] 구체적으로, 예측부는 디코딩하려는 임베딩 벡터(v_i)를 조건으로 하는 노드 조건 디코더(f_{dec})를 이용하여 다음 시점의 최종 측정값($\hat{x}^t$)을 예측할 수 있다. 즉, 예측부는 수학식 3에 따라 다음 시점의 최종 예측 측정값($\hat{x}^t$)을 획득할 수 있다.

수학식 3

[0078] $$\hat{x}^t = f_{dec}(z^t; v_i)$$

[0079] 상술된 인코더, 업데이트 함수, 디코더는 인공 신경망을 이용할 수 있다. 인공 신경망 및 노드 임베딩은 평균 제곱 오차(Mean Square Error, MSE)를 손실 함수로 사용함으로써 학습할 수 있다.

[0081] 도 5는 본 개시의 일 실시예에 따른, 이상 탐지부가 수행하는 동작을 설명하기 위한 도면이다.

[0082] 이상 탐지부는 제2 시점에서의 예측 시계열 데이터(즉, 예측된 측정 값) 및 제2 시점에서의 실제 시계열 데이터(즉, 실제 측정 값)을 이용해서 해당 시점(즉, 제2 시점)에서 실체 측정 값의 이상 유무를 판단할 수 있다.

[0083] 먼저, 이상 탐지부는 제2 시점에서의 예측 시계열 데이터 및 제2 시점에서의 실제 시계열 데이터를 입력으로 이용하는 이상 점수 계산 함수 f_{ano}를 이용하여 해당 시점에서 각 센서 별로 이상 점수를 계산할 수 있다(S510).

[0084] 예로, 이상 탐지부는 수학식 4를 이용하여 해당 시점에서 각 센서 별로 이상 점수를 계산할 수 있다.

수학식 4

[0085] $$s_i^t = f_{ano}(x_i^t, \hat{x}_i^t)$$

[0086] 이상 탐지부는 계산된 이상 점수에 대해 후처리를 수행할 수 있다(S520).

[0087] 구체적으로, 센서 별로 계산된 이상 점수의 통계량이 다른 경우가 많으므로, 이상 탐지부는 제2 시점에서의 예측 시계열 데이터와 제2 시점에서의 실제 시계열 데이터 간의 비교를 위하여 계산된 이상 점수에 대해 후처리(예로, 정규화 등)를 수행할 수 있다.

[0088] 이상 탐지부는 후처리가 수행된 이상 점수를 이용하여 최종 이상 점수를 계산할 수 있다(S530).

[0089] 구체적으로, 이상 탐지부는 최종 이상 점수 함수(f_{select})를 이용하여 해당 시점에서 센서 별 이상 점수를 통합하여 하나의 최종 이상 점수를 계산할 수 있다. 최종 이상 점수 함수는 수학식 5와 같이 구현될 수 있다.

수학식 5

[0090] $$s^t = f_{select}(\boldsymbol{s}^t)$$

[0091] 여기서, 최종 이상 선택을 위한 함수는 해당 시점에서 센서의 이상 점수들의 평균 값 또는 이상 점수들의 최댓값 등을 획득하기 위한 함수로 구현될 수 있다.

[0092] 이상 탐지부는 최종 이상 점수를 이용하여 해당 시점에서의 데이터들의 이상 여부를 탐지할 수 있다(S540).

[0093] 구체적으로, 이상 탐지부는 계산된 최종 이상 점수를 미리 설정된 문턱 값(Threshold)과 비교하여 해당 시점에서의 이상 여부를 탐지 및 판단할 수 있다. 여기서, 사용하는 문턱 값은 학습 데이터를 통해서 획득될 수 있으나 이에 제한되는 것은 아니다. 문턱 값은 경험 및 실험 데이터 등 다양한 방식에 따라 설정될 수 있다.

[0095] 도 6은 본 개시의 일 실시예에 따른, 이상 원인 분석부가 수행하는 동작을 설명하기 위한 도면이다.

[0096] 이상 탐지부에 의해 해당 시점에서의 이상이 탐지되는 경우(즉, 제2 시점에서의 실제 시계열 데이터의 이상이 탐지되는 경우), 이상 원인 분석부는 이상이 발생된 센서를 식별 및 선택할 수 있다(S610).

[0097] 일 예로, 이상 탐지부는 이상의 원인을 식별하기 위하여 이상 점수가 가장 높은 센서를 이상이 발생된 센서로

선택할 수 있다. 또 다른 예로, 사용자에 의해 (경험적으로 또는 실험적으로 이상 점수가 가장 높은 센서가 아닌) 다른 센서가 선택될 수 있다.

[0098] 이상 탐지부는 선택된 이상이 발생한 센서에게 가장 영향을 많이 미치는 센서들의 우선순위를 계산할 수 있다(S620).

[0099] 일 예로, 그래프 구조에 기초하여 이상이 발생한 센서와 관련된 모든 센서에 대한 정보를 확인함으로써 이상이 발생한 센서에게 가장 영향을 많이 미치는 센서(즉, 발생/탐지된 이상에 많은 영향을 미친 센서)의 우선순위가 확인될 수 있다. 다만, 그래프 구조를 이용하여 이상이 발생한 센서와 관련된 모든 센서에 대한 정보를 확인하는 방식은 많은 시간 및 지식을 필요로 할 수 있다. 또 다른 예로, 계층 별 관련성 전파(Layer wise Relevance Propagation, LRP)를 이용하여 적어도 하나의 센서에 대한 우선순위를 설정할 수 있다.

[0100] LRP를 이용하여(즉, 수학식 6에 따라) 이상이 발생한 센서와 이에 영향을 주는 적어도 하나의 센서 간의 관련성을 산출할 수 있다. 장치는 각 적어도 하나의 센서 별로 이상이 발생한 센서와 관련성이 높은 순으로 우선 순위를 결정할 수 있다.

수학식 6

[0101]
$$R_i^{(l)} = \sum_k \frac{x_i^{(l)} w_{ik}^{(l,l+1)}}{\sum_j x_j^{(l)} w_{jk}^{(l,l+1)}} R_k^{(l+1)}$$

[0102] 여기서, $x_j^{(l)}$ 는 l번째 계층에서의 j번째 특징 벡터이고, $w_{ik}^{(l,l+1)}$ 는 l번째 계층에서 $l+1$번째 계층으로 연결되는 가중치 중 l번째 계층 특징 벡터의 i번째 요소에서 $l+1$번째 계층 특징 벡터의 k번째 요소로 가는 가중치를 의미할 수 있다.

[0103] 장치는 마지막 계층의 관련성 점수를 이상이 발생한 것으로 선택된 센서의 예측 결과로서 사용할 수 있다. 수학식 6에 따라 관련성 점수가 산출되면, 최종적으로 입력 데이터의 모양과 같이 시간 창 길이와 노드의 수의 크기가 2차원 행렬 형태로 구현될 수 있다. 장치는 이를 각 노드 별로 합산함으로써 최종적으로 적어도 하나의 노드(즉, 센서) 별로 관련도(즉, 관련성 점수)를 획득할 수 있다.

[0104] 그리고, 이상 탐지부는 이상 원인을 시각화하여 제공할 수 있다(S630).

[0105] 일 예로, 적어도 하나의 노드 별로 획득된 관련도에 기초하여 도 4에 도시된 바와 이상이 탐지된 노드 및 이와 관련된 노드들을 시각화할 수 있다.

[0106] 따라서, 이상 탐지부는 이상이 발생한 센서와 관련된 센서들의 이상 발생의 우선순위를 계산하고, 계산된 우선순위에 기초하여 우선순위 높은 센서를 찾을 수 있도록 도 4와 같이 이상 원인을 시각화하여 제공할 수 있다.

[0108] 도 7은 본 개시의 일 실시예에 따른, 그래프 네트워크를 이용하여 다변량 시계열 데이터 이상을 탐지하는 방법을 설명하기 위한 순서도이다.

[0109] 장치는 복수의 센서 각각으로부터 획득된 제1 시점에서의 제1 시계열 데이터에 대해 전처리를 수행할 수 있다(S710).

[0110] 여기서, 전처리 절차는 제1 시점에서의 제1 시계열 데이터에 대해 정규화(normalization) 또는 패스트 푸리에 변환(fast Fourier transformation) 절차를 포함할 수 있다.

[0111] 구체적으로, 장치는 전처리부를 통해 새로운 시계열 데이터(또는, 측정 값)를 저장하고, 미리 저장된 데이터(또는, 미리 저장된 데이터)를 이용하여 시간 창을 구성한다. 이후에 장치는 학습 데이터의 최댓값과 최솟값으로 새로운 시계열 데이터를 사용하여 Min-Max 정규화 절차를 수행할 수 있다.

[0112] 여기서, 데이터의 저장은 제1 시점에서의 제1 시계열 데이터(즉, 미리 지정된 길이 만큼의 과거 데이터)를 저장할 수 있다. 최근 시점(예로, 제2 시점 또는 그 이후의 시점)에 데이터가 입력될 경우, 장치는 저장된 지 가장

오래된 데이터를 삭제하고 최근 시점에 입력된 데이터를 저장할 수 있다. 데이터의 정규화는 Min-Max 정규화 절차 또는 Fast Fourier Transformation 중의 적어도 하나를 포함할 수 있으나 이에 제한되는 것은 아니다.

[0113] 장치는 전처리된 제1 시계열 데이터에 기초하여 복수의 센서 별로 제2 시점에서의 예측 시계열 데이터를 획득할 수 있다(S720).

[0114] 구체적으로, 장치는 복수의 센서 각각을 복수의 노드로서 설정한 그래프를 생성하고, 전처리된 제1 시계열 데이터를 인코딩하여 복수의 센서 각각에 대응되는 노드의 표현을 획득할 수 있다. 즉, 그래프 구조에서 노드는 하나의 센서에 대응될 수 있다.

[0115] 복수의 센서 각각에 대응되는 노드의 표현 및 그래프에 기초하여, 장치는 복수의 센서 각각에 대응되는 노드의 표현(representation)을 업데이트할 수 있다. 장치는 복수의 센서 각각에 대응되는 노드의 표현으로부터 제2 시점에서의 예측 시계열 데이터를 획득할 수 있다.

[0116] 여기서, 그래프(즉, 복수의 센서에 대응되는 노드로 구성된 그래프 구조)는 복수의 센서 중 제1 센서(예로, 소스 센서)에 대응되는 제1 노드 및 제1 노드와 연결된 제2 센서(예로, 타겟 센서)에 대응되는 제2 노드를 포함할 수 있다.

[0117] 제1 노드의 임베딩 벡터 및 제2 노드의 임베딩 벡터를 통해 제2 노드의 표현 데이터가 제1 노드의 표현에 기초하여 업데이트될 수 있다. 그리고, 제1 센서에 대응되는 제1 노드의 임베딩 벡터에 기초하여 업데이트된 제1 노드의 노드 표현을 디코딩(decoding)함으로써, 제1 센서에 대한 제2 시점에서의 예측 시계열 데이터가 획득될 수 있다.

[0118] 본 개시의 일 예로, 장치는 예측부를 통해 학습된 임베딩 벡터들 간의 코사인 유사도를 수학식 1에 따라 획득할 수 있다. 장치는 각 노드 별로 유사도가 가장 높은 k 개의 노드를 선택하여 그래프 구조를 구성할 수 있다. 이 때, k는 사용자에 의해 결정될 수 있으며 미리 정의될 수도 있다. 장치는 전처리된 데이터를 인코더를 통하여 인코딩을 수행할 수 있다.

[0119] 이 때, 인코더는 다층 퍼셉트론(Multi-Layer Perceptron, MLP) 또는 1차원 합성곱 네트워크 등과 같은 인공 신경망으로 사용될 수 있다.

[0120] 장치는 도 8에 도시된 바와 같이 인코딩된 입력과 그래프 구조를 이용하여 노드의 표현을 업데이트할 수 있다. 즉, 장치는 연결되어 있는 두 노드의 임베딩 벡터를 조건으로 하는 간선 별 노드 업데이트 모듈을 통해 각 노드의 표현을 업데이트 할 수 있다.

[0121] 상술된 바와 같이, 장치는 업데이트된 노드 표현을 노드 조건 디코더를 이용하여 수학식 3과 같이 각 노드 별로 제2 시점에서의 예측 시계열 데이터를 획득할 수 있다. 간선 별 노드 업데이트 모듈 및 노드 조건 디코더는 MLP와 같은 인공 신경망으로 구현될 수 있다.

[0122] 복수의 센서 각각으로부터 획득된 제2 시점에서의 실제 시계열 데이터 및 제2 시점에서의 예측 시계열 데이터에 기초하여, 장치는 복수의 센서 각각으로부터 획득된 제2 시점에서의 실제 시계열 데이터의 이상 여부를 탐지할 수 있다(S730).

[0123] 장치는 복수의 센서 각각으로부터 획득된 제2 시점에서의 실제 시계열 데이터 및 제2 시점에서의 예측 시계열 데이터를 이용하여 복수의 센서 별 이상 점수를 산출할 수 있다. 장치는 복수의 센서 별 이상 점수에 대해 후처리(post processing)을 수행할 수 있다.

[0124] 장치는 후처리된 복수의 센서 별 이상 점수에 기초하여 복수의 센서 별 최종 이상 점수를 산출할 수 있다. 장치는 복수의 센서 별 최종 이상 점수를 통해 복수의 센서 각각으로부터 획득된 제2 시점에서의 실제 시계열 데이터의 이상 여부를 탐지할 수 있다.

[0125] 구체적으로, 장치는 이상 점수 계산 함수에 예측된 측정 값(즉, 제2 시점에서의 예측 시계열 데이터) 및 실제 측정 값(즉, 제2 시점(예로, 현재)에서의 실제시계열 데이터)을 적용하여 각 센서 별로 이상 점수를 산출할 수 있다.

[0126] 여기서, 이상 점수 계산 함수는 학습에 사용된 MSE를 포함할 수 있으나 이에 제한되는 것은 아니다. 장치는 노드 간의 이상 점수 등의 비교를 위하여 후처리를 수행할 수 있다.

[0127] 후처리 방식으로 검증 데이터 셋의 평균과 분산을 사용하는 정규화를 포함할 수 있으나 이에 제한되는 것은 아

니다. 장치는 수학식 5와 같이 최종 이상 선택 함수에 후처리 된 이상 점수를 입력하여 최종적으로 해당 시간에서의 최종 이상 점수를 산출할 수 있다.

[0128] 장치는 최종 이상 선택 함수로서 최댓값 함수 뿐만 아니라 평균이나 가중평균과 같이 센서 별 이상 점수를 이용하여 계산하는 계산식을 사용할 수 있다. 최종적으로 계산된 최종 이상 점수가 미리 설정된 문턱 값보다 클 경우, 장치는 이상이 발생한 것으로 탐지할 수 있다.

[0129] 제2 시점에서의 실제 시계열 데이터의 이상이 탐지됨에 기반하여, 장치는 복수의 센서 중 탐지된 이상과 관련된 특정 센서에 대한 정보를 제공할 수 있다(S740).

[0130] 여기서, 특정 센서에 대한 정보는 복수의 센서 중 특정 센서와 관련된 적어도 하나의 센서 및 특정 센서 간의 관련도를 나타내는 그래프 정보를 포함할 수 있다.

[0131] 본 개시의 일 예로, 장치는 적어도 하나의 센서가 특정 센서에 영향을 주는 정도에 기초하여 적어도 하나의 센서 각각의 우선순위를 산출할 수 있다. 여기서, 적어도 하나의 센서는 적어도 특정 센서에 대응되는 노드에 연결된 적어도 하나의 노드에 대응될 수 있다.

[0132] 장치는 적어도 하나의 센서 각각의 우선순위에 기초하여 적어도 하나의 센서 및 특정 센서 간의 관련도를 산출할 수 있다. 즉, 적어도 하나의 센서 및 특정 센서 간의 관련도는 적어도 하나의 센서가 특정 센서에 미치는 영향의 크기를 나타낼 수 있다.

[0133] 구체적으로, 이상이 발생된 것으로 탐지되면, 장치는 이상 원인 분석부를 통해 이상이 탐지된 센서를 선택한다. 장치는 선택된 센서에 대해 가장 영향을 많이 주는 적어도 하나의 센서를 선택하고, 적어도 하나의 센서에 대한 우선순위를 설정할 수 있다.

[0134] 여기서, 장치는 우선 순위 설정 방식으로서 계층별 관련성 전파를 이용하여 적어도 하나의 센서에 대한 중요도를 수학식 6으로 계산하여 이를 바탕으로 우선 순위를 설정할 수 있다

[0135] 이상, 본 개시에서 설명되는 실시예와 첨부된 도면은 본 개시에 포함되는 기술적 사상의 일부를 예시적으로 설명하는 것에 불과하다. 따라서, 본 명세서에 개시된 실시예들은 본 개시의 기술적 사상을 한정하기 위한 것이 아니라 설명하기 위한 것이므로, 이러한 실시예에 의하여 본 발명의 기술 사상의 범위가 한정되는 것이 아님은 자명하다.

[0136] 본 개시의 명세서 및 도면에 포함된 기술적 사상의 범위 내에서 당업자가 용이하게 유추할 수 있는 변형 예와 구체적인 실시 예는 모두 본 개시의 권리범위에 포함되는 것으로 해석되어야 할 것이다.

부호의 설명

[0137] 100 : 장치

110 : 메모리

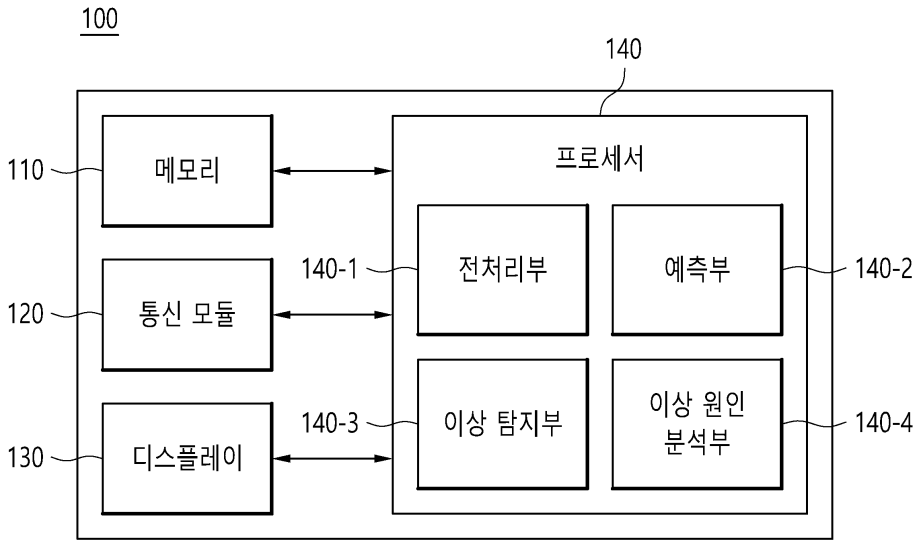
120 : 통신 모듈

130 : 디스플레이

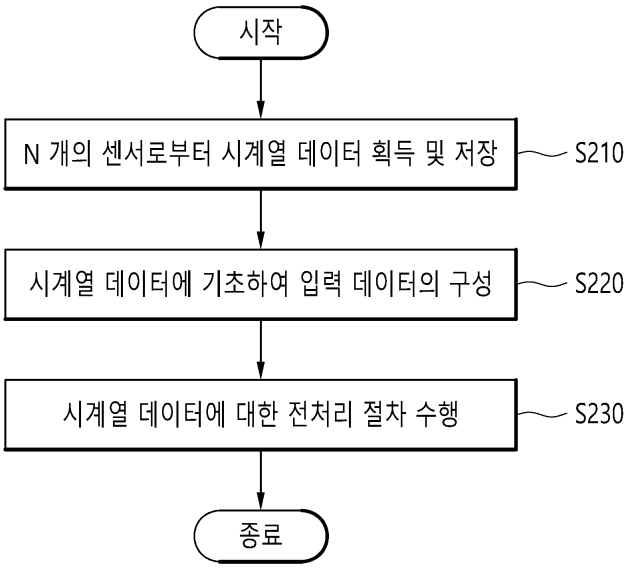
140 : 프로세서

도면

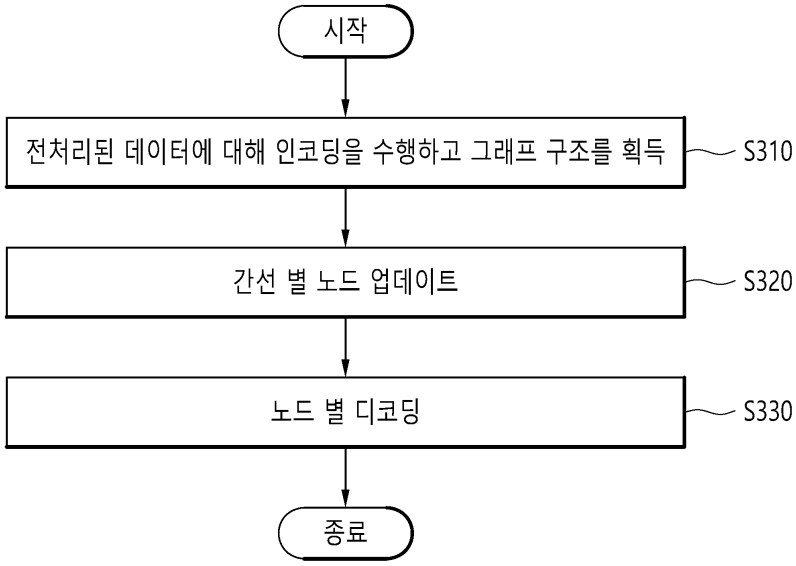
도면1


100
140
110
메모리
프로세서
140-1
전처리부
예측부
140-2
120
통신 모듈
130
디스플레이
140-3
이상 탐지부
이상 원인
분석부
140-4

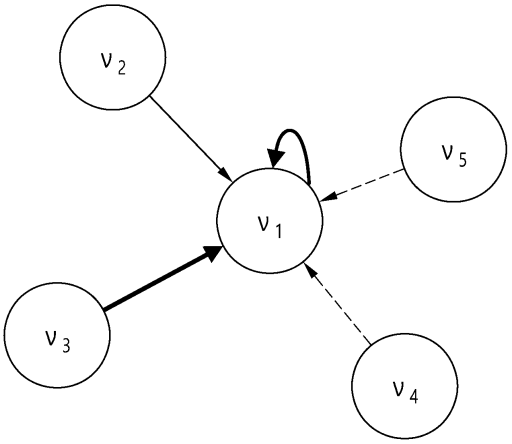
도면2


시작
N 개의 센서로부터 시계열 데이터 획득 및 저장
S210
시계열 데이터에 기초하여 입력 데이터의 구성
S220
시계열 데이터에 대한 전처리 절차 수행
S230
종료

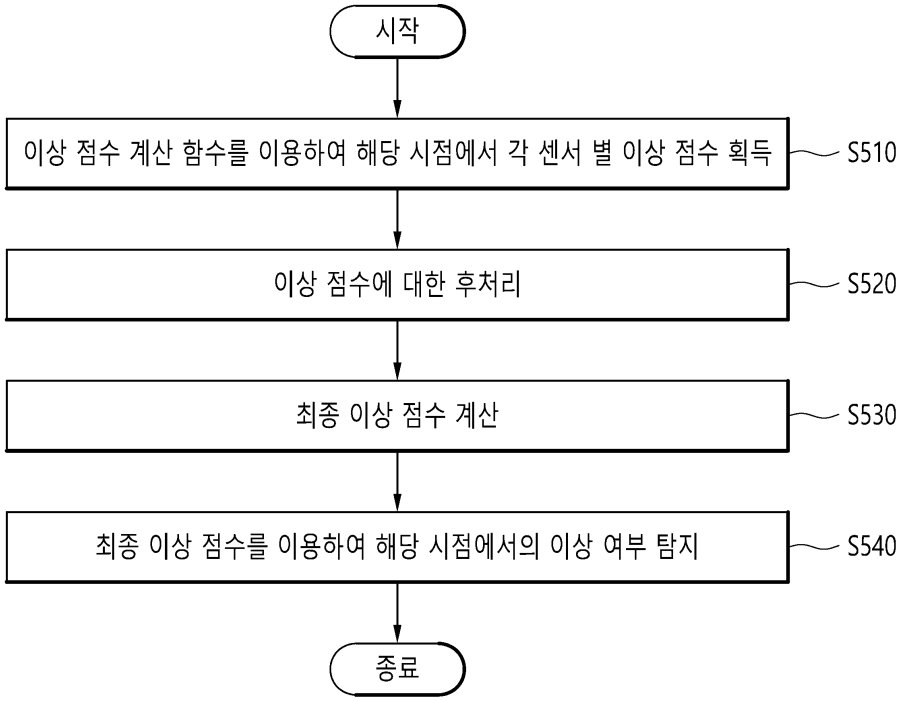
도면3


시작
전처리된 데이터에 대해 인코딩을 수행하고 그래프 구조를 획득
S310
간선 별 노드 업데이트
S320
노드 별 디코딩
S330
종료

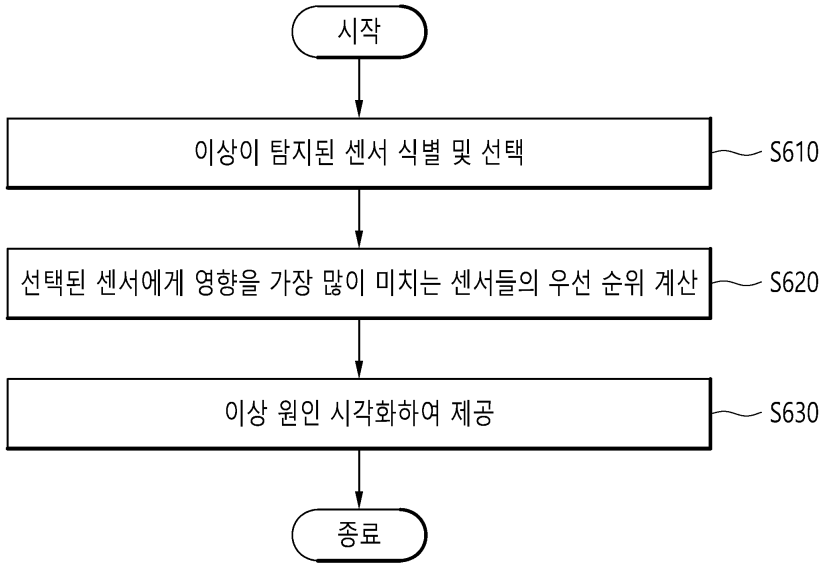
도면4


v2
v5
v1
v3
v4

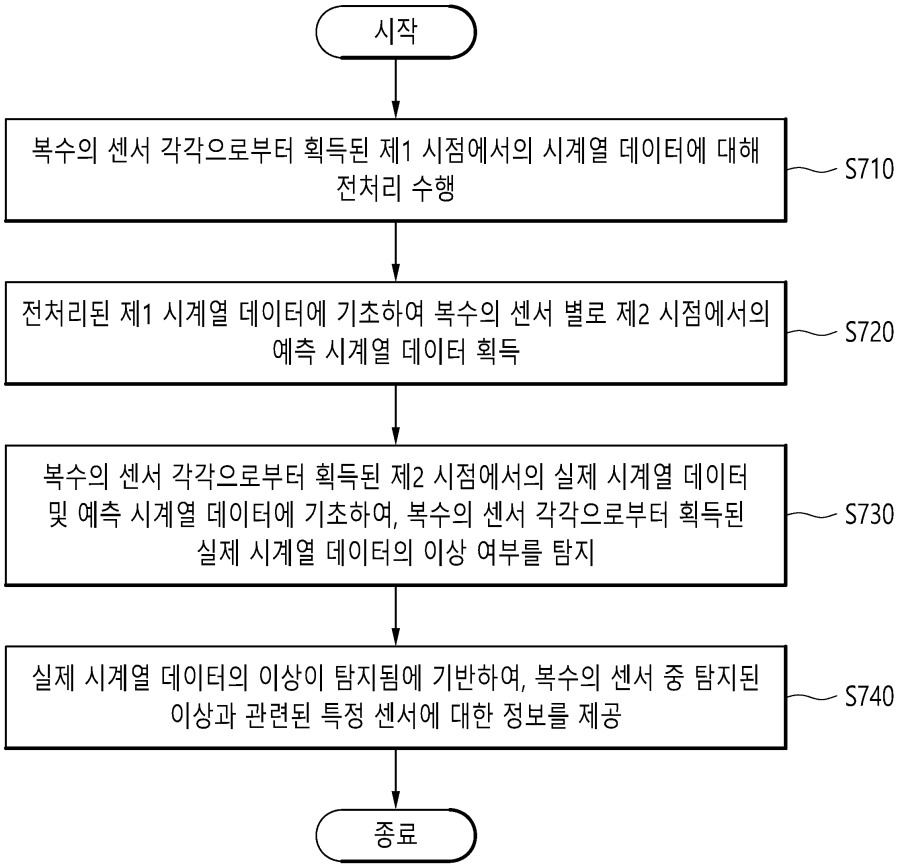
도면5


시작
이상 점수 계산 함수를 이용하여 해당 시점에서 각 센서 별 이상 점수 획득
S510
이상 점수에 대한 후처리
S520
최종 이상 점수 계산
S530
최종 이상 점수를 이용하여 해당 시점에서의 이상 여부 탐지
S540
종료

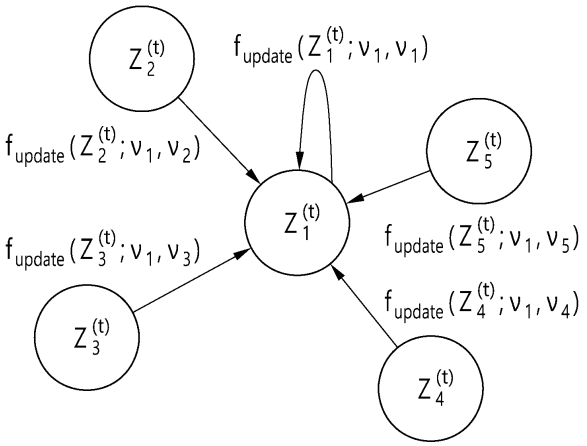
도면6


시작
이상이 탐지된 센서 식별 및 선택
S610
선택된 센서에게 영향을 가장 많이 미치는 센서들의 우선 순위 계산
S620
이상 원인 시각화하여 제공
S630
종료

도면7


시작
복수의 센서 각각으로부터 획득된 제1 시점에서의 시계열 데이터에 대해 전처리 수행
S710
전처리된 제1 시계열 데이터에 기초하여 복수의 센서 별로 제2 시점에서의 예측 시계열 데이터 획득
S720
복수의 센서 각각으로부터 획득된 제2 시점에서의 실제 시계열 데이터 및 예측 시계열 데이터에 기초하여, 복수의 센서 각각으로부터 획득된 실제 시계열 데이터의 이상 여부를 탐지
S730
실제 시계열 데이터의 이상이 탐지됨에 기반하여, 복수의 센서 중 탐지된 이상과 관련된 특정 센서에 대한 정보를 제공
S740
종료

도면8


$Z_2^{(t)}$
$f_{update}(Z_1^{(t)};v_1,v_1)$
$f_{update}(Z_2^{(t)};v_1,v_2)$
$Z_5^{(t)}$
$Z_1^{(t)}$
$f_{update}(Z_3^{(t)};v_1,v_3)$
$f_{update}(Z_5^{(t)};v_1,v_5)$
$f_{update}(Z_4^{(t)};v_1,v_4)$
$Z_3^{(t)}$
$Z_4^{(t)}$