(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2025-0005789
(43) 공개일자 2025년01월10일

(51) 국제특허분류(Int. Cl.)
G06N 3/045 (2023.01) *G06N 3/04* (2023.01)
G06N 3/0442 (2023.01)
(52) CPC특허분류
G06N 3/045 (2023.01)
G06N 3/0442 (2023.01)
(21) 출원번호 10-2023-0086035
(22) 출원일자 2023년07월03일
심사청구일자 2023년07월03일

(71) 출원인
고려대학교 산학협력단
서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)
(72) 발명자
백준걸
서울특별시 강동구 고덕로 333, 102동 3302호
문지원
서울특별시 송파구 올림픽로 435, 221동 1002호
송승환
전라북도 전주시 완산구 서원로 289, 109동 1002호
(74) 대리인
김등용

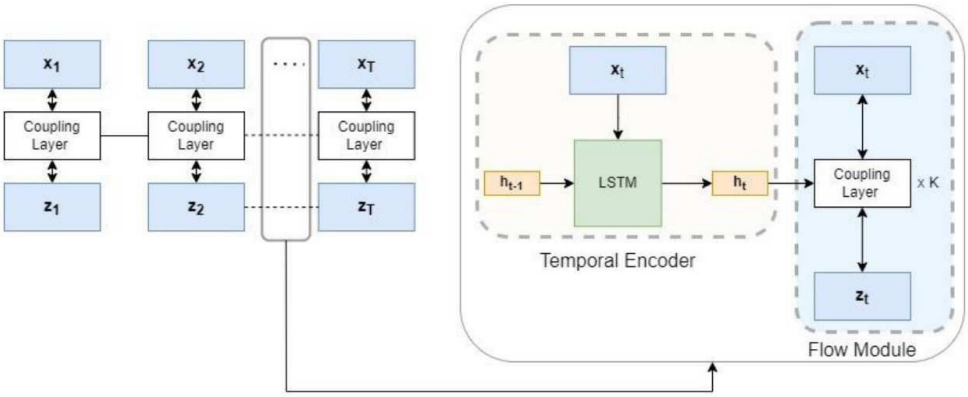
전체 청구항 수 : 총 6 항

(54) 발명의 명칭 **NF를 활용한 다변량 시계열 이상 탐지 장치 및 방법**

(57) 요 약

NF(Normalizing Flow)를 활용한 다변량 시계열 데이터의 이상 탐지 장치 및 방법이 개시된다. 상기 이상 탐지 방법은 적어도 프로세서(processor)를 포함하는 컴퓨팅 장치에 의해 수행되는 다변량 시계열(Multivariate Time Series) 데이터의 이상 탐지 방법으로써, 학습 데이터를 획득하는 단계, 및 상기 학습 데이터를 이용하여 신경망 모델을 학습시켜 탐지 모델을 생성하는 단계를 포함하고, 상기 학습 데이터는 다변량 시계열 데이터이고, 상기 신경망 모델은 입력 데이터에 대한 히든 스테이트(hidden state)를 출력하는 템퍼럴 인코더(Temporal Encoder) 모델과 상기 히든 스테이트를 조건부로 하는 상기 입력 데이터의 분포를 학습하는 NF(Normalizing Flow) 모델로 구성된다.

대 표 도 - 도5



(52) CPC특허분류
G06N 3/049 (2023.01)

이 발명을 지원한 국가연구개발사업
과제고유번호 1711157319
과제번호 2022R1A2C2004457
부처명 과학기술정보통신부
과제관리(전문)기관명 한국연구재단
연구사업명 개인기초연구(과기정통부)
연구과제명 멀티 모달 스트리밍 데이터의 잠재 공간에서의 표현학습 기반 크로스 모달 네트워크 개발
기 여 율 1/1
과제수행기관명 고려대학교
연구기간 2023.03.01 ~ 2024.02.29

명 세 서

청구범위

청구항 1

적어도 프로세서(processor)를 포함하는 컴퓨팅 장치에 의해 수행되는 다변량 시계열(Multivariate Time Series) 데이터의 이상 탐지 방법에 있어서,

학습 데이터를 획득하는 단계; 및

상기 학습 데이터를 이용하여 신경망 모델을 학습시켜 탐지 모델을 생성하는 단계를 포함하고,

상기 학습 데이터는 다변량 시계열 데이터이고,

상기 신경망 모델은 입력 데이터에 대한 히든 스테이트(hidden state)를 출력하는 템퍼럴 인코더(Temporal Encoder) 모델과 상기 히든 스테이트를 조건부로 하는 상기 입력 데이터의 분포를 학습하는 NF(Normalizing Flow) 모델로 구성된,

이상 탐지 방법.

청구항 2

제1항에 있어서,

상기 템퍼럴 인코더 모델은 양방향 LSTM(Bidirectional Long Short Term Memory)을 이용하는,

이상 탐지 방법.

청구항 3

제1항에 있어서,

다변량 시계열 데이터인 타겟 데이터를 수신하는 단계; 및

상기 탐지 모델을 이용하여 상기 타겟 데이터의 이상 여부를 판단하는 단계를 더 포함하는,

이상 탐지 방법.

청구항 4

제3항에 있어서,

상기 이상 여부를 판단하는 단계는, 상기 탐지 모델에 포함된 NF 모델의 출력이 미리 정해지 임계치 보다 작은 경우, 이상이 발생한 것으로 판단하는,

이상 탐지 방법.

청구항 5

제1항에 있어서,

상기 탐지 모델을 생성하는 단계는, 입력 데이터에 대한 가능도(likelihood)가 최대화되도록 상기 신경망 모델을 학습하는,

이상 탐지 방법.

청구항 6

제1항에 있어서,

상기 탐지 모델을 생성하는 단계는, 수학식에 의해 정의되는 손실 함수(loss function, L)를 최소화하도록 학습되고,

상기 수학식은 $$\mathcal{L} = \frac{1}{T}\sum_{t=1}^{T} -\log p_X(x_t|h_t)$$ 이고,

상기 p_x는 상기 NF 모델의 입력 공간(input space)의 밀도(density)이고, 상기 x_t는 타임 스텝(time step) t에서의 입력 데이터이고, 상기 h_t는 타임 스텝 t에서 상기 템퍼럴 인코더 모델의 출력인 히든 스테이트이고, 상기 T는 타임 윈도우(time window)의 크기인,

이상 탐지 방법.

발명의 설명

기 술 분 야

[0001] 본 발명은 제조 공정 등에서 발생하는 불균형 다변량 시계열 데이터(Multivariate Time Series Data)에 대해 효과적으로 이상(Anomaly)을 탐지하기 위한 방법에 관한 것으로, 특히 NF(Normalizing Flow) 기반의 네트워크를 이용하여 변수 간의 상관 관계와 시간 상의 의존 관계를 모두 반영해 이상을 탐지하는 방법에 관한 것이다.

배 경 기 술

[0002] 확률적으로 정의되는 이상(Anomaly)이란 정규성의 개념에서 크게 벗어나는 관측치들을 의미한다. 즉, 이상을 탐지한다는 것은 대부분의 데이터로부터 멀리 떨어진 비정상적인 관측치를 식별하는 연구 분야이다. 특히, 제조 공정에서의 이상 탐지는 실제 산업 현장의 생산 수율과 직결된다. 따라서, 신속하고 정확한 이상 탐지가 이루어져야 작업 중단과 품질 저하로부터 발생할 수 있는 막대한 경제적 손실을 예방할 수 있다.

[0003] 최근 다수의 센서로부터 수집되는 다변량 시계열 데이터의 이상 탐지에 대한 연구의 필요성이 증가하고 있으며, 이러한 데이터에 대한 신속하고 정확한 이상 탐지는 제조 공정뿐만 아니라 다양한 산업을 보다 생산적으로 만들 수 있다. 본 발명에서는 제조 공정에 NF(Normalizing Flow)라는 생성(generative) 모델을 사용한 이상 탐지 모델을 통해 생산성과 수율을 높이는 방법을 제안한다.

[0004] 다수의 센서들로부터 수집되는 다변량 시계열 데이터에 대한 이상 탐지는 두 가지의 이유로 어려움을 겪는다. 첫째, 고차원 변수 간의 상관 관계(correlation)뿐만 아니라 각각의 시계열 데이터의 복잡한 시간 의존성(temporal dependency)까지 반영하여야 한다. 제조 공정에 사용되는 대부분의 기존 방법들은 센서들 간의 상관 관계와 시간 상의 의존 관계를 제대로 학습하지 못한다. 두 가지 특성을 모두 반영하는 것은 모델의 복잡도가 증가하기 때문에 실질적으로 적용하기에 어려움이 따른다. 또한, 최근 두 가지 특성을 모두 반영하기 위한 2 단계로 구성된 접근법이 많이 연구되고 있으나, 두 모델을 개별적으로 훈련하여 국소 최적에 빠지는 문제가 발생하기도 한다. 둘째, 제조 공정 데이터는 클래스 불균형(class imbalance) 문제가 자주 발생한다. 공정에서 수집되는 데이터는 이상 현상의 빈도가 낮고, 이상 관측치에 대한 라벨링이 제대로 되어 있지 않은 경우가 대부분이기 때문이다. 따라서 제조 공정에서 실질적으로 사용할 수 있으며 앞서 언급된 두 가지 어려움을 극복할 수 있는, 모델 복잡도를 줄인 엔드 투 엔드(end-to-end) 방식의 비지도 다변량 시계열 이상 탐지 방법론이 필요하다.

[0005] 이러한 문제들을 해결하기 위해 표현 학습(representation learning) 기반 비지도(unsupervised) 이상 탐지 방법론이 많이 연구되었다. 대표적으로, GAN(Generative Adversarial Networks), 그리고 VAE(Variational AutoEncoder)를 이용한 모델들이 높은 차원을 갖는 데이터의 분포를 모델링하기 위해 사용되었다. 그러나, GAN 기반 모델은 데이터를 직접 생성하는 확률적인(stochastic) 절차만 정의하고, VAE는 잠재 벡터(latent vector) z에 대한 확률 분포(probability distribution)를 다루기 쉬운 확률 분포에 대략적으로만 근사한다는 한계가 존재한다.

[0006] NF(Normalizing Flow)는 연속된 가역적인(invertible) 변환으로 구성되어 데이터의 분포를 명시적으로(explicitly) 학습하므로, 가능도(likelihood)를 정확하게 계산할 수 있다는 장점이 존재한다. 이는 가능도의 크고 작음을 기준으로 이상을 탐지하는 확률론적 이상 탐지에 적합할 수 있다.

[0007] 본 발명에서는 제조 공정에서의 두 가지 실질적인 문제를 극복하기 위해 NF를 활용한 다변량 시계열 이상 탐지 방법론을 제안한다. NF를 사용해 변수 간의 관계를 반영하고 데이터의 분포를 학습하며, 데이터의 시간 의존성을 반영하기 위해 LSTM(Long Short Term Memory) 기반의 모델들을 사용해 장기 의존성을 반영한다. 이를 통해 변수 간의 상관 관계와 시간 상의 의존 관계를 모두 반영해 이상을 탐지할 수 있다. 또한, 본 발명은 제조 공정

데이터의 클래스 불균형 문제를 해결하기 위해 다수 클래스인 정상(혹은 대부분이 정상인) 데이터만으로 모델을 학습한다. 따라서, 제조 공정에서 데이터 불균형 문제 해결과 동시에 높은 이상 탐지 성능을 제공하고자 한다. 이는 제조 공정에서의 높은 생산성으로 이어질 수 있다.

선행기술문헌

특허문헌

[0008] (특허문헌 0001) 대한민국 공개특허 제2021-0132549호 (2021.11.04. 공개)

(특허문헌 0002) 대한민국 공개특허 제2023-0055609호 (2023.04.26. 공개)

(특허문헌 0003) 대한민국 등록특허 제2428632호 (2022.08.03. 공고)

(특허문헌 0004) 대한민국 공개특허 제2022-0110113호 (2022.08.05. 공개)

발명의 내용

해결하려는 과제

[0009] 본 발명이 이루고자 하는 기술적인 과제는 클래스 불균형이 존재하는 제조 공정 데이터에서 다변량 시계열 데이터의 시간적, 공간적 특성을 반영하기 위해 NF와 LSTM 기반의 네트워크를 활용하는 것이다. 제안하는 기법에 의할 경우, 제조 공정에서 높은 정확도로 이상을 탐지할 수 있어, 생산성과 수율을 높일 수 있다.

과제의 해결 수단

[0010] 본 발명의 일 실시예에 따른 이상 탐지 방법은 적어도 프로세서(processor)를 포함하는 컴퓨팅 장치에 의해 수행되는 다변량 시계열(Multivariate Time Series) 데이터의 이상 탐지 방법으로써, 학습 데이터를 획득하는 단계, 및 상기 학습 데이터를 이용하여 신경망 모델을 학습시켜 탐지 모델을 생성하는 단계를 포함하고, 상기 학습 데이터는 다변량 시계열 데이터이고, 상기 신경망 모델은 입력 데이터에 대한 히든 스테이트(hidden state)를 출력하는 템퍼럴 인코더(Temporal Encoder) 모델과 상기 히든 스테이트를 조건부로 하는 상기 입력 데이터의 분포를 학습하는 NF(Normalizing Flow) 모델로 구성된다.

발명의 효과

[0011] 본 발명의 실시예에 따른 다변량 시계열 이상 탐지 방법론에 의할 경우, 제조 공정에서 다수의 센서들로부터 수집된 클래스 불균형이 존재하는 다변량 시계열 데이터에 대한 이상 탐지에 적용될 수 있다.

[0012] 실제 산업에 적용될 경우, 기존 모델 대비 높은 이상 탐지 성능을 보여 주어 생산성 및 수율 향상을 기대할 수 있다. 본 발명 모델에 대한 학습을 선행적으로 완료된 상태에서 실제 공정에 실시간으로 적용할 시 정확도 높은 이상 탐지가 가능하다.

도면의 간단한 설명

[0013] 본 발명의 상세한 설명에서 인용되는 도면을 보다 충분히 이해하기 위하여 각 도면의 상세한 설명이 제공된다.

도 1은 NF(Normalizing Flow)의 구조를 도시한다.

도 2는 LSTM의 구조를 도시한다.

도 3은 양방향 LSTM의 구조를 도시한다.

도 4는 Input Attention Encoder의 구조를 도시한다.

도 5는 본 발명의 일 실시예에 따른 이상 탐지 방법의 개요도이다.

도 6은 본 발명에서 제안하는 기법의 AUROC를 나타내는 그래프이다.

도 7은 본 발명에 대한 성능 평가 결과로써, 특정 임계값을 설정해 이상을 탐지한 결과를 도시한다.

도 8은 본 발명의 성능과 다른 모델들에 의한 성능의 비교 결과를 도시한다.

도 9는 본 발명의 일 실시예에 따른 이상 탐지 방법을 설명하기 위한 흐름도이다.

발명을 실시하기 위한 구체적인 내용

[0014] 본 명세서에 개시되어 있는 본 발명의 개념에 따른 실시예들에 대해서 특정한 구조적 또는 기능적 설명들은 단지 본 발명의 개념에 따른 실시예들을 설명하기 위한 목적으로 예시된 것으로서, 본 발명의 개념에 따른 실시예들은 다양한 형태들로 실시될 수 있으며 본 명세서에 설명된 실시예들에 한정되지 않는다.

[0015] 본 발명의 개념에 따른 실시예들은 다양한 변경들을 가할 수 있고 여러 가지 형태들을 가질 수 있으므로 실시예들을 도면에 예시하고 본 명세서에서 상세하게 설명하고자 한다. 그러나, 이는 본 발명의 개념에 따른 실시예들을 특정한 개시 형태들에 대해 한정하려는 것이 아니며, 본 발명의 사상 및 기술 범위에 포함되는 모든 변경, 균등물, 또는 대체물을 포함한다.

[0016] 제1 또는 제2 등의 용어는 다양한 구성 요소들을 설명하는데 사용될 수 있지만, 상기 구성 요소들은 상기 용어들에 의해 한정되어서는 안 된다. 상기 용어들은 하나의 구성 요소를 다른 구성 요소로부터 구별하는 목적으로만, 예컨대 본 발명의 개념에 따른 권리 범위로부터 벗어나지 않은 채, 제1 구성 요소는 제2 구성 요소로 명명될 수 있고 유사하게 제2 구성 요소는 제1 구성 요소로도 명명될 수 있다.

[0017] 어떤 구성 요소가 다른 구성 요소에 "연결되어" 있다거나 "접속되어" 있다고 언급된 때에는, 그 다른 구성 요소에 직접적으로 연결되어 있거나 또는 접속되어 있을 수도 있지만, 중간에 다른 구성 요소가 존재할 수도 있다고 이해되어야 할 것이다. 반면에, 어떤 구성 요소가 다른 구성 요소에 "직접 연결되어" 있다거나 "직접 접속되어" 있다고 언급된 때에는 중간에 다른 구성 요소가 존재하지 않는 것으로 이해되어야 할 것이다. 구성 요소들 간의 관계를 설명하는 다른 표현들, 즉 "~사이에"와 "바로 ~사이에" 또는 "~에 이웃하는"과 "~에 직접 이웃하는" 등도 마찬가지로 해석되어야 한다.

[0018] 본 명세서에서 사용한 용어는 단지 특정한 실시예를 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.

[0019] 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 의미로 해석되지 않는다.

[0020] 이하, 본 명세서에 첨부된 도면들을 참조하여 본 발명의 실시예들을 상세히 설명한다. 그러나, 특허출원의 범위가 이러한 실시예들에 의해 제한되거나 한정되는 것은 아니다. 각 도면에 제시된 동일한 참조 부호는 동일한 부재를 나타낸다.

[0021] 우선, NF(Normalizing Flow)와 LSTM(Long Short Term Memory)에 대해 설명한다.

[0022] **1. NF(Normalizing Flow)**

[0023] NF(Normalizing Flow)는 심층 생성 모델(Deep Generative Model) 계열로, 도 1과 같은 구조를 갖는다.

[0024] NF는 입력 데이터 x가 역함수가 존재하는 플로우 레이어(Flow layer) f(x)를 연속적으로 거쳐 잠재 공간에서의 기본 분포(underlying distribution) z를 학습하고 역변환(invertible)된 플로우 레이어 $f^{-1}(z)$를 통해 데이터를 복원(x′)할 수 있다. 주요한 특징은 입력(input)과 출력(output)의 차원(dimension)이 같다는 것으로, 입력 공간(input space) $X \in \mathbb{R}^D$의 밀도(density) p_X가 간단한 분포를 따르는 출력 공간(output space) $Z \in \mathbb{R}^D$의 밀도 p_Z (예: isotropic Gaussian)로 변환되도록 $\mathbb{R}^D$에서 $\mathbb{R}^D$로 매핑(mapping)된다. 이러한 매핑 $f : X \mapsto Z$은 일련의 전단사(bijection) 또는 역변환(invertible) 함수로 구성된다. 이를 변수 변환(change of variables)이

라고 하는데, 변수 변환 공식에 기반하여 $X = f^{-1}(Z)$로 표현된다. 변수 변환된 x의 밀도 $p_X(x)$를 수학식 1과 같이 나타낼 수 있다.

[0025] [수학식 1]

[0026] $$p_X(x) = p_Z(z)\left|\det\left(\frac{\partial f(x)}{\partial x}\right)\right|$$

[0027] 수학식 1에서 $\frac{\partial f(x)}{\partial x}$ 는 x에 대한 야코비안(jacobian)을 의미한다. 야코비안 행렬(jacobian matrix)은 벡터 f(x), x에 대한 일차 편미분 행렬이며, x와 z의 확률 밀도 함수(probability density function)의 관계가 야코비안 행렬식(jacobian determinant) 만큼의 비율을 갖는다. 각 레이어(layer)마다 야코비안 행렬식을 계산하는 것은 많은 비용이 들기 때문에, NF는 다음과 같은 두 가지 조건을 만족하여야 한다. 첫째, f는 가역적이어야 한다. 둘째, f에 대한 야코비안 행렬식 계산이 쉬어야 한다. 위의 두 조건을 만족하지 않으면 NF의 학습이 제대로 이루어지지 않는다.

[0028] RealNVP는 야코비안 행렬식의 연산을 간단히 하기 위해 어파인 커플링 레이어(affine coupling layer)라는 플로우 네트워크(flow network) 구성 방법을 제안하였다(Dinh, L., Sohl-Dickstein, J., & Bengio, S. (2016) Density estimation using real nvp arXiv preprint arXiv:1605.08803.). 이는 입력의 일부를 변환하지 않고, 변환되지 않은 변수들의 함수를 통해 나머지 변수를 변환한다. 이를 수학식 2와 같이 나타낼 수 있다.

[0029] [수학식 2]

[0030] $$\begin{cases} y^{1:d} = b \odot x^{1:d} \\ y^{d+1:D} = (1-b)\left(x^{d+1:D} + t(b \odot x^{1:d})\right) \\ \qquad \odot exp\left(s(b \odot x^{1:d})\right) \end{cases}$$

[0031] 수학식 2에서 $x^{1:d}$는 변환되지 않은 변수, $x^{d+1:D}$는 변환되는 변수를 의미한다. 또한, $\odot$는 성분별 곱(element wise product)이고, $s(\cdot)$는 스케일링(scaling), $t(\cdot)$는 변환 함수(translation function)로 각각 선형적(linear)이지 않은 복잡한 함수이다. b는 이진 마스킹(binary masking)으로 역전파 시에 예측되어야 할 마스크 처리된(masked) 부분을 의미하며 수학식 3을 통해 업데이트 된다.

[0032] [수학식 3]

[0033] $$x_{masked} = (x_{masked} + t(x_{1:d})) \odot \exp\left(s(x_{1:d})\right)$$

[0034] 커플링 레이어(Coupling layer)에서는 마스크 처리된 입력(masked input)에 대한 어파인 변환(affine transform)을 수행하기 위해 s(·), t(·) 함수를 학습한다.

[0035] 선형적(Linear)이지 않은 함수 f를 모델링하기 위해서 $X \mapsto Y_1 \mapsto Y_2 \mapsto \cdots \mapsto Y_{K-1} \mapsto Z$ 가 되도록 다수의 커플링 레이어가 구성된다. 또한, 변경되지 않는 차원 $(e.g.\, x^{1:d})$이 항상 동일하지 않도록 셔플링(shuffling)이 수행된다. 변수 변환 공식(change of variable formula)에 입각해 데이터 포인트 x에 대한 확률 밀도 함수를 수학식 4와 같이 나타낼 수 있다.

[0036] [수학식 4]

[0037] $$\log p_X(x) = \log p_Z(z) + \log\left|\det\left(\frac{\partial z}{\partial x}\right)\right|$$
$$= \log p_Z(z) + \sum_{i=1}^{K} \log\left|\det\left(\partial y_i / \partial y_{i-1}\right)\right|$$

[0038] 어파인 커플링 레이어(Affine coupling layer)에 의해 RealNVP의 야코비안은 수학식 5와 같은 하삼각행렬(lower triangular matrix) 형태로 변환이 되고, 야코비안 연산은 수학식 6과 같이 스케일링 함수 출력(scaling function output)의 합으로 매우 간단해 진다.

[0039] [수학식 5]

$$\frac{\partial y_i}{\partial y_{i-1}} = \begin{bmatrix} \mathbb{I} & 0 \\ \frac{\partial y_i^{d+1:D}}{\partial y_{i-1}^{1:d}} & diag\left(\exp\left(s(y_{i-1}^{1:d})\right)\right) \end{bmatrix}$$

[0040]

[0041] [수학식 6]

$$\log|\det(\partial y_i/\partial y_{i-1})| = \log|\exp(sum(s_i(y_{i-1}^{1:d}))|$$

[0042]

[0043] NF 모델의 훈련은 일반적으로 훈련 데이터의 가능도를 최대화하는 방향으로 학습된다. 즉, T 기간 동안의 데이터에 대한 음의 로그 가능도(negative log likelihood)를 최소화 하도록 학습하게 된다. 최종적으로 NF의 손실 함수(loss function)는 수학식 7과 같다.

[0044] [수학식 7]

$$\mathcal{L} = \frac{1}{T}\sum_{t=1}^{T} -\log p_X(x;\theta)$$

[0045]

[0046] ## 2. LSTM(Long Short Term Memory)

[0047] ### 2.1 Vanilla LSTM

[0048] LSTM은 RNN(Recurrent Neural Network)에서의 기울기 소실(vanishing gradient) 문제를 해결하기 위해 고안된 모델이다. LSTM의 구조는 도 2와 같다.

[0049] 매 타임 스텝(time step)마다 2개의 스테이트(state, 상태): 셀 스테이트(sell state(C_t)), 히든 스테이트(hidden state(h_t))를 유지하며 3 개의 게이트(gate): 망각 게이트(forget gate(f)), 입력 게이트(input gate(i)), 출력 게이트(output gate(o))을 통해 계산이 이루어 진다. t 시점의 입력 벡터가 $x_t \in \mathbb{R}^n$일 때, 히든 스테이트 $h_t \in \mathbb{R}^m$는 수학식 8 내지 수학식 12를 통해 도출된다

[0050] [수학식 8]

[0051] $$i_t = sigm(W_i[h_{t-1}; x_t] + b_i)$$

[0052] [수학식 9]

[0053] $$f_t = sigm(W_f[h_{t-1}; x_t] + b_f)$$

[0054] [수학식 10]

[0055] $$o_t = sigm(W_o[h_{t-1}; x_t] + b_o)$$

[0056] [수학식 11]

[0057] $$C_t = f_t \odot C_{t-1} + i_t \odot tan\,h(W_C[h_{t-1}; x_t] + b_C)$$

[0058] [수학식 12]

[0059] $$h_t = o_t \circ \tanh(C_t)$$

[0060] 수학식 8 내지 수학식 11의 $[h_{t-1}; x_t] \in \mathbb{R}^{m+n}$은 이전 시점의 히든 스테이트인 h_{t-1}과 현재 입력(input)인 x_t를 컨케트네이트(concatenate) 해준 값이며 $W_i, W_f, W_o, W_C \in \mathbb{R}^{m\times(m+n)}$과 $b_i, b_f, b_o, b_C \in \mathbb{R}^m$는 학습 파라미터를 의미한다. 수학식 8에서 입력 게이트(input gate)는 셀 스테이트를 업데이트하는데 사용된다. t 시점의 입력과 t-1 시점의 히든 스테이트가 시그모이드(sigmoid) 함수로 전달되어 0과 1 사이의 값으로 변환하고 새로 업데이트 될 값을 결정한다. 수학식 9의 망각 게이트(forget gate)는 0에서 1 사이의 숫자를 제공한다. 1은 해당 정보를

모두 유지하는 것을 의미하며, 0은 모두 잊는 것을 의미한다. 수학식 10의 출력 게이트(output gate)는 이전 입력에 대한 정보를 갖고 있고 다음 히든 스테이트를 결정한다. 수학식 11의 $\tanh(W_C[h_{t-1}; x_t] + b_C)$는 입력 변조(input modulation)로 -1에서 1 사이의 값을 갖으며, 현재 정보를 셀 스테이트에 얼마나 더할 것인지를 결정하는 역할을 수행한다. 최종적으로 t 시점의 히든 스테이트는 수학식 12와 같이 계산된다.

[0061] ## 2.2 Bidirectional LSTM

[0062] 양방향 LSTM(Bidirectional LSTM)은 양방향성을 갖는 LSTM으로 도 3과 같은 구조로 이루어 진다.

[0063] 도 3에서의 파란색 화살표는 전방향(forward) LSTM을 의미하며, 초록색 화살표는 역방향(backward) LSTM을 의미한다. 이후 각각 전방향의 t번째 셀과 역방향의 (N-t)번째 셀을 연결(concatenate)하여 최종 출력을 얻게 된다. 양방향 LSTM은 시퀀스 데이터에 대한 정보를 양방향으로 추출할 수 있어 일반적으로 Vanilla LSTM 보다 성능이 우수하다는 특징이 있다.

[0064] ## 2.3 Input Attention Encoder

[0065] Qin et al.(Qin, Y., Song, D., Chen, H., Cheng, W., Jiang, G., & Cottrell, G. (2017) A dual-stage attention-based recurrent neural network for time series prediction arXiv preprint arXiv:1704.02971.)은 서로 관련이 있는 입력 변수(input feature)를 적응적으로(adaptively) 선택할 수 있는 입력 어텐션 기반 인코더(Input Attention based Encoder)를 제안하였다. 입력 어텐션 인코더(Input Attention Encoder)의 구조는 도 4에 도시되어 있다.

[0066] k번째 차원의 $\mathbf{x^k} = (x_1^k, x_2^k, \cdots, x_T^k)^\intercal \in \mathbb{R}^T$가 주어졌을 때, 수학식 13, 수학식 14와 같이 Encoder LSTM 유닛의 h_{t-1}과 C_{t-1}를 참조하고, 멀티레이어 퍼셉트론(multilayer perceptron)을 통해서 어텐션 메커니즘(attention mechanism)을 구성할 수 있다.

[0067] [수학식 13]

[0068] $$e_t^k = v_e^\intercal \tanh(W_e[h_{t-1}; C_{t-1}] + U_e \mathrm{x^k})$$

[0069] [수학식 14]

[0070] $$a_t^k = \frac{\exp(e_t^k)}{\sum_{i=1}^{n} \exp(e_t^i)}$$

[0071] 수학식 13에서 $v_e \in \mathbb{R}^T, W_e \in \mathbb{R}^{T \times 2m}$, 그리고 $U_e \in \mathbb{R}^{T \times T}$는 각각 학습 파라미터이다. 수학식 14에서 a_t^k는 시점 t에서의 k번째 입력 변수 중요도를 계산하는 attention weight로, 합이 1이 될 수 있도록 e_t^k에 소프트맥스(softmax) 함수가 적용된다. 이러한 attention weight를 사용해 수학식 15에서 변수 중요도가 반영된 입력 값을 반환할 수 있고, 수학식 16을 통해 t 시점의 히든 스테이트가 업데이트된다.

[0072] [수학식 15]

[0073] $$\widetilde{x_t} = (a_t^1 x_t^1, a_t^2 x_t^2, \cdots, a_t^n x_t^n)$$

[0074] [수학식 16]

[0075] $$h_t = f_1(h_{t-1}, \widetilde{x_t})$$

[0076] 수학식 16에서의 f_1은 수학식 8 내지 수학식 12에 따라 계산되는 LSTM 유닛이다.

[0077] 이하에서는, 본 발명에서 제안하는 이상 탐지 기법을 상세하게 설명한다. 본 발명은 변수 간의 관계와 시간 의존성을 학습할 수 있는 NF 기반의 다변량 시계열 이상 탐지 기법을 제안한다. 제안 방법의 전체 구조는 도 5와 같다. 제안 기법은 다음과 같이 3 단계로 구성된다. (1) Temporal Encoder: 데이터가 LSTM을 거쳐 시계열을 요약하는 정보를 담고 있는 히든 스테이트를 반환한다. (2) Flow Module: 도출된 히든 스테이트를 조건부로 하는 데이터의 분포를 NF을 통해 모델링한다. (3) Training: NF의 출력(output)으로 도출된 분포에 대한 네거티브 로그 가능도(negative log likelihoot)를 손실 함수로써 학습한다. 각 방법은 유기적인 순서로 학습이 진행될 수 있다.

[0078] 1. Temporal Encoder

[0079] 본 발명에서는 시간 의존성을 반영하기 위해 템퍼럴 인코더(Temporal Encoder) 구조를 사용한다. 템퍼럴 인코더(Temporal Encoder)에서는 LSTM 기반 모델들을 사용한다. 본 발명에서는 히든 스테이트를 도출하기 위해 도 2의 Vanilla LSTM, 도 3의 양방향 LSTM, 그리고 도 4의 Input Attention Encoder 중 적어도 하나를 사용한다. Vanilla LSTM은 가장 기본적인 LSTM의 형태로, 단기 상태를 고려하는 동시에 장기 상태에 대한 정보를 공급받을 수 있는 모델이다. 양방향 LSTM은 Vanilla LSTM을 양방향으로 겹쳐 놓은 구조로 순차적으로 입력된 데이터에 대해서 이전 데이터와의 관계뿐만 아니라 이후 데이터와의 관계까지 학습한다. Input Attention은 변수(feature)에 대해 어텐션(attention)을 주어 중요한 변수를 중점적으로 학습할 수 있다.

[0080] 즉, 제안 방법은 시간 의존성뿐만 아니라 공간적(spatial)인 정보까지 추가로 고려하여 히든 스테이트를 추출할 수 있는 여러 방법을 고려한다. Vanilla LSTM과 양방향 LSTM은 시간 상의 정보를 중점적으로 반영하는 반면, Input Attention Encoder는 시간 상의 정보를 반영하는 동시에 중요한 변수를 중점적으로 학습할 수 있다. 변수 시점 t에서 D개의 차원을 갖는 데이터 $x_t \in \mathbb{R}^D$로부터 t 시점의 히든 스테이트 h_t는 수학식 17과 같이 계산된다.

[0081] [수학식 17]

[0082] $$h_t = LSTM(x_t, h_{t-1})$$

[0083] 수학식 17에서 h_t는 t 시점까지의 시계열을 요약하는 정보를 담고 있으며, Flow Module의 커플링 레이어(Coupling layer)에 입력된다.

[0084] 2. Flow Module

[0085] 데이터의 분포를 학습하기 위해, Flow Module을 사용해 역함수가 존재하는 함수를 학습한다. 제안 방법은 Temporal Encoder에 사용된 LSTM과 함께 도 5에서 볼 수 있듯, K 개의 Flow Module(RealNVP 혹은 MAF)을 결합하여 분포를 모델링한다. LSTM으로부터 나온 히든 스테이트를 입력 데이터 x와 연결(concatenate)한 후, Flow Module의 커플링 레이어에 입력한다. Flow Module의 커플링 레이어에서는 파라미터의 공유를 통해 변수 간의 관계를 파악할 수 있다. 커플링 레이어는 수학식 2와 같이 스케일(scaling)과 변환 함수(translation function)로 구성되어 있으며, 수학식 18과 같이 스케일과 변환 함수 s(·), t(·)에 입력 데이터 x와 히든 스테이트 h가 결합된 채 입력된다.

[0086] [수학식 18]

[0087] $$s(\text{concat}(x^{1:d}, h))$$
$$t(\text{concat}(x^{1:d}, h))$$

[0088] RealNVP에서는 어파인 커플링 레이어(affine coupling layer)가, MAF에서는 MADE(Masked Autoencoder for Distribution Estimation, Germain, M., Gregor, K., Murray, I., & Larochelle, H. (2015) Made: Masked autoencoder for distribution estimation. In International conference on machine learning (pp. 881-889). PMLR.)가 커플링 레이어로서 야코비안을 하삼각행렬 형태로 구성하고, 복잡한 연산을 간단히 할 수 있다. 최종적으로 t 시점에서의 로그 밀도(log density)는 수학식 19와 같다.

[0089] [수학식 19]

[0090] $$\log p_X(x_t|h_t) = \log q_X\big(f(x_t|h_t)\big) + \log\left|\det\left(\frac{\partial f(x_t|h_t)}{\partial x_t}\right)\right|$$

[0091] 3. Training

[0092] 제안 방법은 템퍼럴 인코더(Temporal Encoder)와 플로우 모듈(Flow Module)을 거쳐 유기적으로 학습되며, 최종 손실 함수는 수학식 7에서 수학식 20과 같이 변경된다.

[0093] [수학식 20]

[0094] $$\mathcal{L} = \frac{1}{T}\sum_{t=1}^{T} -\log p_X(x_t|h_t)$$

[0095] 제안하는 탐지 모델은 데이터 x에 대한 가능도를 최대화한다. 손실을 최소화하는 관점에서 가능도에 음수를 취하여, 음의 로그 가능도(negative log likelihood)를 손실 함수로써 학습한다. 또한, T 길이의 윈도우(window) 내의 모든 데이터에 대해서 연산해야 하기 때문에 시점의 길이인 T로 나누어 손실의 평균을 계산한다. 이를 기준으로 제안 방법의 학습을 진행한다.

[0096] 본 발명의 성능 평가 지표는 AUROC(Area Under Receiver Operating Characteristic Curve)이다. AUC는 ROC curve의 밑면적으로 모델의 임계값을 연속적으로 바꾸면서 측정한 FPR(False Positive Rate)과 TPR(True Positive Rate)의 변화를 나타낸 것이다. 가능한 모든 임계값에 대해 산출되기 때문에 임계값에 독립적인 성능 평가 지표로 활용될 수 있다. 본 발명을 10번 반복 수행했을 때의 평균 AUC는 0.9250으로 1에 가까운 매우 우수한 성능을 나타냈으며 이에 AUROC는 도 6에 도시되어 있다.

[0097] 본 발명에 대한 성능 평가는 임계값을 설정하지 않고 이루어 질 수 있지만, 특정 임계값을 설정해 이상을 탐지한 결과는 도 7에 도시되어 있다. 한 가지 케이스를 제외하고는 탐지한 이상 윈도우(anomaly window)가 실제 이상 시점을 대부분 정확하게 반영함을 확인할 수 있다.

[0098] 데이터 기반 이상 탐지에 자주 사용되는 다른 모델들과 본 발명을 비교한 결과는 도 8에 도시되어 있다. 두 번째로 성능이 우수한 Deep-SVDD 대비 본 발명의 성능이 5% 이상 향상되었음을 확인할 수 있다.

[0099] 도 9는 본 발명의 일 실시예에 따른 이상 탐지 방법을 설명하기 위한 흐름도이다. 이상 탐지 방법은 다변량 시계열 이상 탐지 방법 등으로 명명될 수 있으며, 적어도 프로세서(processor) 및/또는 메모리(memory)를 포함하는 컴퓨팅 장치에 의해 수행될 수 있다. 따라서, 이상 탐지 방법을 이루는 단계들 중 적어도 일부는 컴퓨팅 장치에 포함되는 프로세서의 동작으로 이해될 수도 있다. 컴퓨팅 장치는 PC(Personal Computer), 서버(server), 랩탑 컴퓨터, 태블릿 PC 등을 포함할 수 있다. 이상 탐지 방법을 설명함에 있어, 앞선 기재와 중복되는 내용에 관하여는 그 구체적인 기재를 생략하기로 한다.

[0100] 우선, 학습 데이터가 획득된다(S110). 학습 데이터는 유무선 통신망을 통하여 외부 디바이스로부터 수신되거나, USB 메모리 장치와 같은 저장 장치로부터 입출력 인터페이스를 통해 수신되거나, 및/또는 이상 탐지 방법을 수행하는 컴퓨팅 장치에 미리 저장되어 있을 수 있다.

[0101] 학습 데이터는 제조 공정 중에서 발생하는 다변량 시계열 데이터를 의미할 수 있다. 즉, 다차원의 시계열 데이터를 의미할 수 있다. 또한, 학습 데이터는 클래스(class)가 정상, 즉 정상 데이터만을 의미할 수 있다. 클래스 불균형 문제를 해소하기 위하여, 정상 데이터만을 학습하기 때문이다.

[0102] 소정의 신경망 모델을 학습함으로써 탐지 모델이 생성된다(S120). 신경망 모델은 템퍼럴 인코더(Temporal Encoder) 모델과 NF(Normalizing Flow) 모델의 결합으로 구성될 수 있고, 템퍼럴 인코더는 Vanilla LSTM, 양방향(Bidirectional) LSTM, Input Attention Encoder 중 어느 하나에 기반할 수 있다.

[0103] 보다 구체적으로, 템퍼럴 인코더는 시계열을 요약하는 정보를 담고 있는 히든 스테이트를 출력(또는 반환)한다. NF 모델은 힌든 스테이트를 조건부로 하는 입력 데이터의 분포를 학습하게 된다. 이때, 템퍼럴 인코더 모델과 NF 모델은 입력 데이터에 대한 가능도(likelihood)를 최대화 하도록 학습된다. 손실을 최소화하는 관점에서 가능도에 음수를 취하여, 네거티브 로그 가능도를 손실 함수로써 이용할 수 있다. 즉, 최종 손실 함수는 상술한 수학식 20과 같다.

[0104] 타겟 데이터가 수신된다(S130). 타겟 데이터는 이상 여부를 탐지하기 위한 대상으로써, 다변량 시계열 데이터일 수 있다. 타겟 데이터는 소정의 유무선 통신망을 통하여 외부 디바이스로부터 실시간으로 또는 비실시간으로 수신될 수 있다. 실시예에 따라, 타겟 데이터는 이상 탐지 방법을 수행하는 컴퓨팅 장치에 포함되는 저장부에 미리 저장되어 있을 수도 있다.

[0105] 학습이 완료된 탐지 모델을 이용하여 타겟 데이터에 대한 이상 여부가 판단될 수 있다(S140). 일 실시예에 따르면, 탐지 모델의 출력(예컨대, 가능도)가 미리 정해진 임계치에 미만되는 경우, 이상이 발생한 것으로 판단할 수 있다. 탐지 모델은 입력 데이터에 대한 가능도를 최대화하고, 학습에 사용된 정상 데이터에 대한 로그 가능도를 최대화하는 학습 프로세스를 통해 표준 정규 분포의 형태로 잠재 공간 상에 임베딩하도록 생성되었기 때문에, 임계치 보다 작은 가능도를 갖는 입력 데이터에는 이상이 있다고 판단될 수 있다.

[0106] 이상에서 설명된 장치는 하드웨어 구성 요소, 소프트웨어 구성 요소, 및/또는 하드웨어 구성 요소 및 소프트웨어 구성 요소의 집합으로 구현될 수 있다. 예를 들어, 실시예들에서 설명된 장치 및 구성 요소는, 예를 들어, 프로세서, 콘트롤러, ALU(Arithmetic Logic Unit), 디지털 신호 프로세서(Digital Signal Processor), 마이크

로컴퓨터, FPA(Field Programmable array), PLU(Programmable Logic Unit), 마이크로프로세서, 또는 명령(instruction)을 실행하고 응답할 수 있는 다른 어떠한 장치와 같이, 하나 이상의 범용 컴퓨터 또는 특수 목적 컴퓨터를 이용하여 구현될 수 있다. 처리 장치는 운영 체제(Operation System, OS) 및 상기 운영 체제 상에서 수행되는 하나 이상의 소프트웨어 애플리케이션을 수행할 수 있다. 또한, 처리 장치는 소프트웨어의 실행에 응답하여, 데이터를 접근, 저장, 조작, 처리 및 생성할 수도 있다. 이해의 편의를 위하여, 처리 장치는 하나가 사용되는 것으로 설명된 경우도 있지만, 해당 기술 분야에서 통상의 지식을 가진 자는, 처리 장치가 복수 개의 처리 요소(Processing Element) 및/또는 복수 유형의 처리 요소를 포함할 수 있음을 알 수 있다. 예를 들어, 처리 장치는 복수 개의 프로세서 또는 하나의 프로세서 및 하나의 콘트롤러를 포함할 수 있다. 또한, 병렬 프로세서(Parallel Processor)와 같은, 다른 처리 구성(Processing Configuration)도 가능하다.

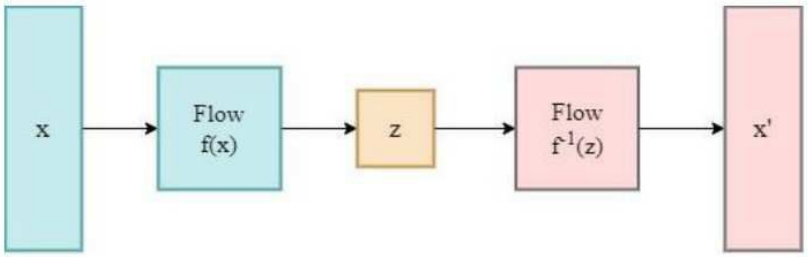
[0107] 소프트웨어는 컴퓨터 프로그램(Computer Program), 코드(Code), 명령(Instruction), 또는 이들 중 하나 이상의 조합을 포함할 수 있으며, 원하는 대로 동작하도록 처리 장치를 구성하거나 독립적으로 또는 결합적으로(Collectively) 처리 장치를 명령할 수 있다. 소프트웨어 및/또는 데이터는, 처리 장치에 의하여 해석되거나 처리 장치에 명령 또는 데이터를 제공하기 위하여, 어떤 유형의 기계, 구성 요소(Component), 물리적 장치, 가상 장치(Virtual Equipment), 컴퓨터 저장 매체 또는 장치, 또는 전송되는 신호 파(Signal Wave)에 영구적으로, 또는 일시적으로 구체화(Embody)될 수 있다. 소프트웨어는 네트워크로 연결된 컴퓨터 시스템 상에 분산되어서, 분산된 방법으로 저장되거나 실행될 수도 있다. 소프트웨어 및 데이터는 하나 이상의 컴퓨터 판독 가능 기록 매체에 저장될 수 있다.

[0108] 실시예에 따른 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 상기 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 상기 매체에 기록되는 프로그램 명령은 실시예를 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(Magnetic Media), CD-ROM, DVD와 같은 광기록 매체(Optical Media), 플롭티컬 디스크(Floptical Disk)와 같은 자기-광 매체(Magneto-optical Media), 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다. 상기된 하드웨어 장치는 실시예의 동작을 수행하기 위해 하나 이상의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.

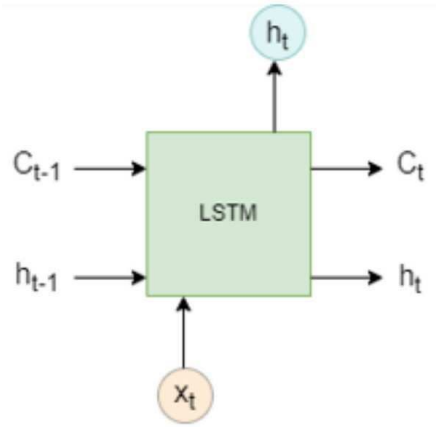
[0109] 본 발명은 도면에 도시된 실시예를 참고로 설명되었으나 이는 예시적인 것에 불과하며, 본 기술 분야의 통상의 지식을 가진 자라면 이로부터 다양한 변형 및 균등한 타 실시예가 가능하다는 점을 이해할 것이다. 예를 들어, 설명된 기술들이 설명된 방법과 다른 순서로 수행되거나, 및/또는 설명된 시스템, 구조, 장치, 회로 등의 구성 요소들이 설명된 방법과 다른 형태로 결합 또는 조합되거나, 다른 구성 요소 또는 균등물에 의하여 대치되거나 치환되더라도 적절한 결과가 달성될 수 있다. 따라서, 본 발명의 진정한 기술적 보호 범위는 첨부된 등록청구범위의 기술적 사상에 의해 정해져야 할 것이다.

도면

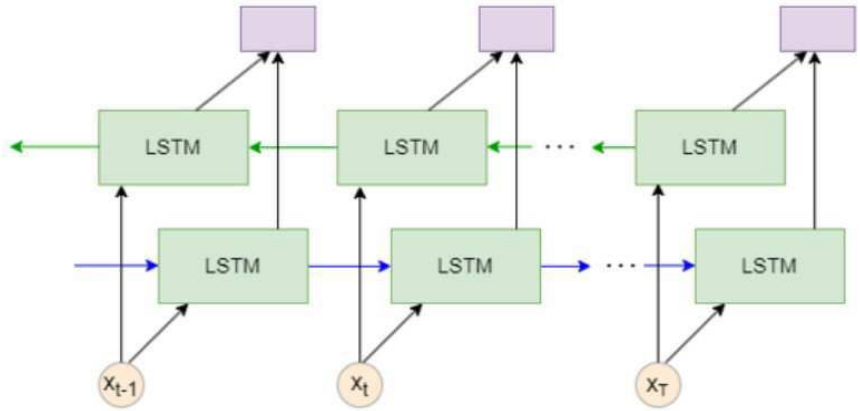
도면1



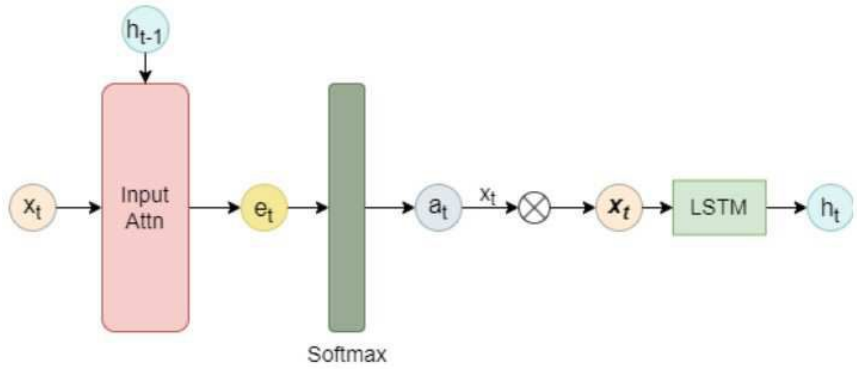
도면2


h_t
C_{t-1}
LSTM
C_t
h_{t-1}
h_t
x_t

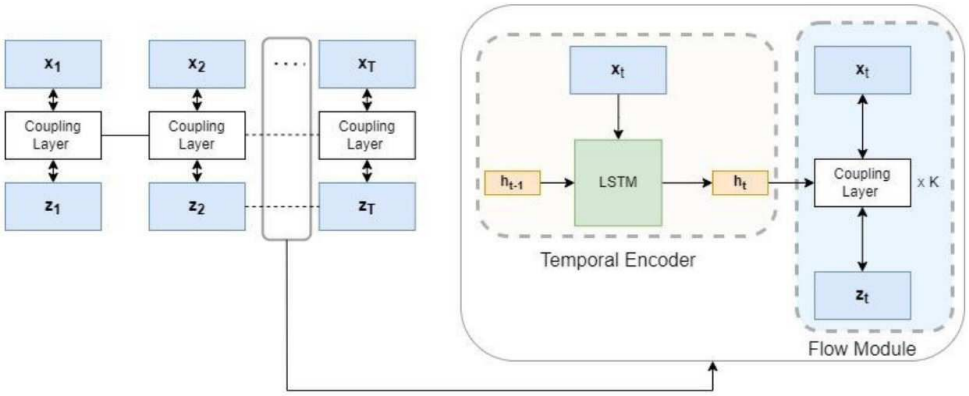
도면3


LSTM
LSTM
LSTM
LSTM
LSTM
LSTM
x_{t-1}
x_t
x_T

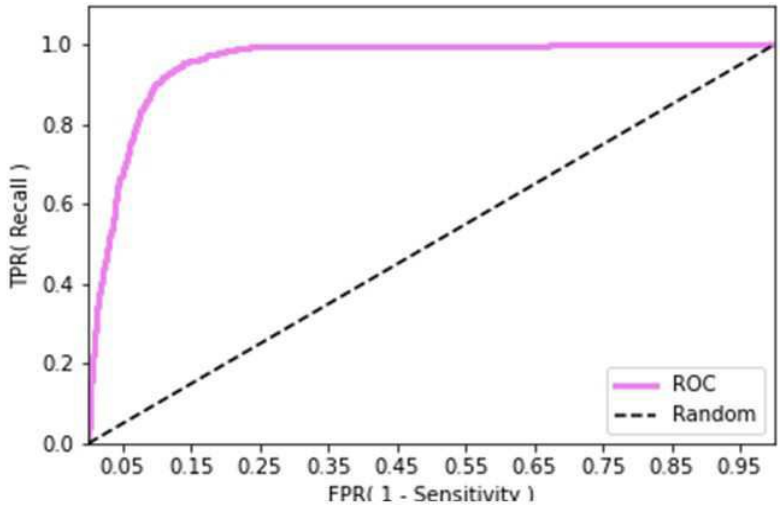
도면4


h_{t-1}
x_t
Input Attn
e_t
Softmax
a_t
x_t
x_t
LSTM
h_t

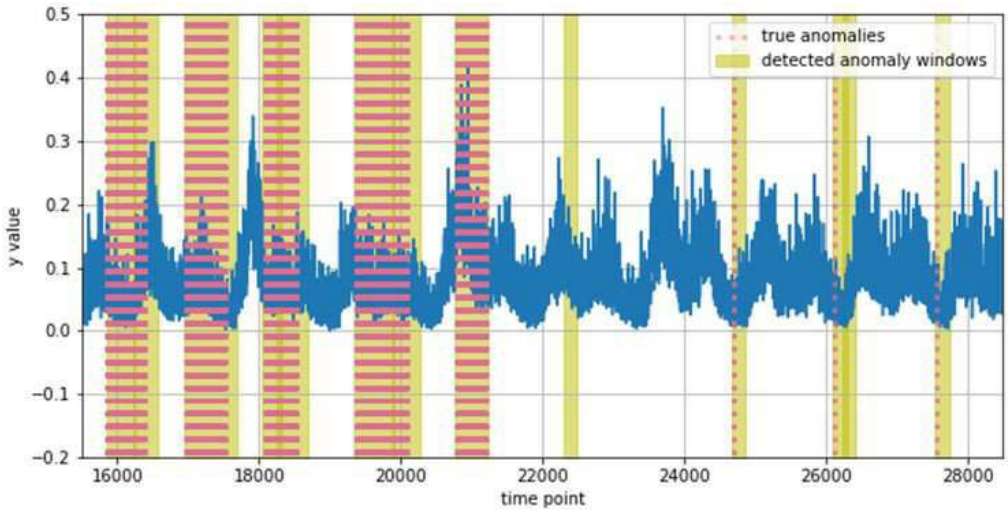
도면5


x1
x2
xT
Coupling Layer
Coupling Layer
Coupling Layer
z1
z2
zT
xt
ht-1
LSTM
ht
Temporal Encoder
xt
Coupling Layer
x K
zt
Flow Module

도면6


1.0
0.8
0.6
0.4
0.2
0.0
TPR(Recall)
0.05
0.15
0.25
0.35
0.45
0.55
0.65
0.75
0.85
0.95
FPR(1 - Sensitivity)
ROC
Random

도면7


0.5
0.4
0.3
0.2
0.1
0.0
−0.1
−0.2
y value
16000
18000
20000
22000
24000
26000
28000
time point
true anomalies
detected anomaly windows

도면8

Method	Parameter #	Training Time	AUC
LSTM Autoencoder	990,465	4558s	0.8587 (0.0087)
MAD-GAN	277,539	190s	0.7906 (0.0208)
Deep-SVDD	57,952	7s	0.8787 (0.0262)
DAGMM	923	14s	0.8731 (0.0790)
RealNVP only	365,052	89s	0.6274 (0.0001)
MAF only	182,844	49s	0.6817 (0.0938)
Proposed Method	387,932	136s	**0.9250 (0.0360)**

도면9

