



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0039729
(43) 공개일자 2024년03월27일

(51) 국제특허분류(Int. Cl.)
G06N 3/08 (2023.01) G06N 20/10 (2019.01)
G06N 3/04 (2023.01)
(52) CPC특허분류
G06N 3/08 (2023.01)
G06N 20/10 (2021.08)
(21) 출원번호 10-2022-0118417
(22) 출원일자 2022년09월20일
심사청구일자 2022년09월20일

(71) 출원인
고려대학교 산학협력단
서울특별시 성북구 안암로 145, 고려대학교 (안암
동5가)
(72) 발명자
백준걸
서울특별시 강동구 고덕로 333, 102동 3302호
황우영
경기도 성남시 분당구 수내로 74, 109동 501호
(74) 대리인
김동용

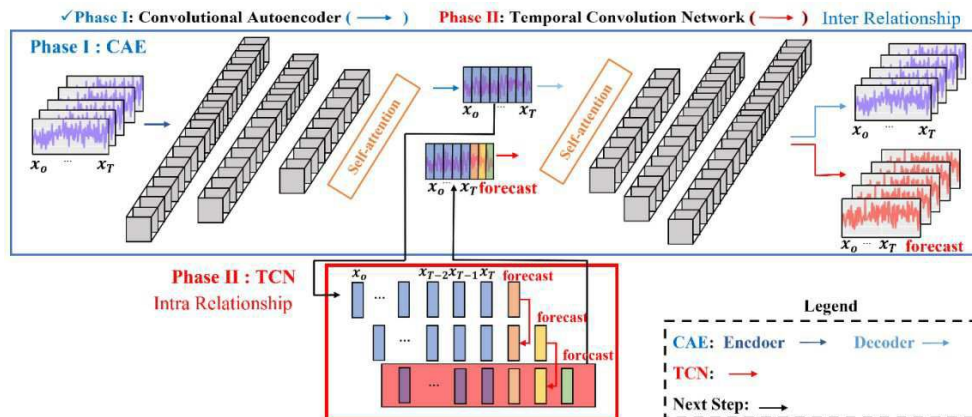
전체 청구항 수 : 총 8 항

(54) 발명의 명칭 SACAE와 TCN을 이용한 다변량 대 다변량 시계열 데이터 예측 장치 및 방법

(57) 요약

SACAE(Self-Attention based Convolutional Autoencoder)와 TCN(Temporal Convolutional Network)을 이용한 다변량 대 다변량 시계열 데이터 예측 모델 생성 방법 및 장치와 이를 이용한 예측 방법이 개시된다. 상기 시계열 데이터 예측 모델 생성 방법은 컴퓨팅 장치에 의해 수행되고, 제1 학습 데이터를 이용하여 SACAE(Self-Attention based Convolutional Autoencoder)를 학습하는 단계, 및 제2 학습 데이터를 이용하여 TCN(Temporal Convolutional Network)을 학습하는 단계를 포함하고, 상기 예측 모델은 상기 SACAE와 상기 TCN을 포함하고, 상기 TCN은 상기 SACAE의 인코더의 출력인 잠재 벡터를 입력으로 수신하여 잠재 벡터의 미래 예측값인 예측 잠재 벡터를 출력하고, 상기 SACAE의 디코더는 상기 예측 잠재 벡터를 입력으로 수신하여 재구성하는 것을 특징으로 한다.

대표도 - 도5



(52) CPC특허분류

G06N 3/045 (2023.01)

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 표현 학습 기반 크로스 모달 네트워크 개발

크 개발

기여율	1/2
과제수행기관명	고려대학교
연구기간	2022.03.01 ~ 2025.02.28

이 발명을 지원한 국가연구개발사업

과제고유번호	1425159221
과제번호	S2837675
부처명	중소벤처기업부
과제관리(전문)기관명	중소기업기술정보진흥원
연구사업명	중소기업기술혁신개발
연구과제명	스마트 팩토리 솔루션의 글로벌 진출을 위한 클라우드 기반의 인공지능 제조 자산

성능 관리 시스템 개발

기여율	1/2
과제수행기관명	고려대학교
연구기간	2020.07.16 ~ 2024.07.15

명세서

청구범위

청구항 1

컴퓨팅 장치에 의해 수행되는 시계열 데이터의 예측 모델 생성 방법에 있어서,

제1 학습 데이터를 이용하여 SACAE(Self-Attention based Convolutional Autoencoder)를 학습하는 단계; 및

제2 학습 데이터를 이용하여 TCN(Temporal Convolutional Network)을 학습하는 단계를 포함하고,

상기 예측 모델은 상기 SACAE와 상기 TCN을 포함하고,

상기 TCN은 상기 SACAE의 인코더의 출력인 잠재 벡터를 입력으로 수신하여 잠재 벡터의 미래 예측값인 예측 잠재 벡터를 출력하고,

상기 SACAE의 디코더는 상기 예측 잠재 벡터를 입력으로 수신하여 재구성하는,

예측 모델 생성 방법.

청구항 2

제1항에 있어서,

상기 SACAE는 상기 인코더와 상기 디코더에 셀프 어텐션(Self-Attention)이 삽입된 CAE(Convolutional Autoencoder)인,

예측 모델 생성 방법.

청구항 3

제1항에 있어서,

상기 SACAE는 재구성 오류가 작아지는 방향으로 학습이 진행되고,

상기 TCN은 예측 오차가 작아지는 방향으로 학습이 진행되는,

예측 모델 생성 방법.

청구항 4

제1항에 있어서,

상기 시계열 데이터는 제조 공정에서 발생하는 다변량 시계열(Multivariate Time Series, MTS) 데이터인,

예측 모델 생성 방법.

청구항 5

컴퓨팅 장치에 의해 수행되는 시계열 데이터의 예측 방법에 있어서,

예측 모델에 포함되는 SACAE(Self-Attention based Convolutional Autoencoder)의 인코더를 이용하여 타겟 데이터에 대한 잠재 벡터를 생성하는 단계;

상기 예측 모델에 포함되는 TCN(Temporal Convolutional Network)을 이용하여 상기 잠재 벡터에 대한 미래 예측값인 예측 잠재 벡터를 생성하는 단계; 및

상기 SACAE의 디코더를 이용하여 상기 예측 잠재 벡터를 디코딩하는 단계를 포함하는 예측 방법.

청구항 6

제5항에 있어서,

상기 SACAE는 상기 인코더와 상기 디코더에 셀프 어텐션(Self-Attention)이 삽입된 CAE(Convolutional Autoencoder)인,

예측 방법.

청구항 7

제5항에 있어서,

상기 타겟 데이터는 제조 공정에서 발생하는 다변량 시계열(Multivariate Time Series, MTS) 데이터인,

예측 방법.

청구항 8

제6항에 있어서,

상기 셀프 어텐션은 상기 인코더의 마지막 레이어와 상기 디코더의 첫번째 레이어에 삽입되는,

예측 방법.

발명의 설명

기술 분야

[0001] 본 발명은 다변량 시계열 데이터의 예측에 관한 것으로, 특히 SACAE(SELF-ATTENTION BASED CONVOLUTIONAL AUTOENCODER)와 TCN(TEMPORAL CONVOLUTIONAL NETWORK) 알고리즘을 이용하여 다변량 시계열 데이터를 효율적으로 예측하는 방법 및 장치에 관한 것이다.

배경 기술

[0002] SACAE(Self-Attention based Convolutional Autoencoder)는 비지도 학습 방식으로 입력의 노이즈(noise)를 무시하고 변수 간의 관계를 학습시킴으로써 다변량 시계열 데이터의 표현을 학습하는 알고리즘이다. 입력을 그대로 복원하고 변수 간의 관계를 학습하는 데에 그 목적이 있다.

[0003] TCN(Temporal Convolutional Network)는 CNN(Convolutional Neural Network) 기반 예측 알고리즘으로 과거 시계열 데이터의 패턴 혹은 개형의 모양을 학습해 다음 시점의 값을 예측하는 알고리즘이다. 미래 시점을 예측하는 데에 그 목적이 있다.

[0004] 다변량 대 다변량 시계열 예측 문제는 미래 시간을 예측하면서 변수와 변수 간의 비선형적인 관계(Inter-relationship)와 각 단변량 시계열 내의 시간적 관계(Intra-relationship)를 모두 통합하여 예측을 수행하는 작업이다.

[0005] 그러나, 저차원 데이터에 적합한 딥러닝 모델은 다변량 시계열 데이터 고유의 고차원 관계를 모델링하는 데 충분하지 않다. 이에, 본 발명에서는 다변량 시계열 데이터가 가지고 있는 비선형적인 관계(Inter-relationship)와 시간적 관계(Intra-relationship)를 모두 고려하여 다변량 시계열 예측을 수행하는 방법과 장치를 제안하고자 한다.

선행기술문헌

특허문헌

[0006] (특허문헌 0001) 대한민국 등록특허 제2038703호 (2019.11.26. 공고)
(특허문헌 0002) 대한민국 등록특허 제2276801호 (2021.07.14. 공고)
(특허문헌 0003) 대한민국 공개특허 제2022-0028724호 (2022.03.08. 공개)

발명의 내용

해결하려는 과제

- [0007] 본 발명이 이루고자 하는 기술적인 과제는 SACAE와 TCN을 이용하여 공정 신호와 같은 다변량 시계열 데이터의 미래값을 효과적으로 예측하는 장치와 방법을 제공하는 것이다.

과제의 해결 수단

- [0008] 본 발명의 일 실시예에 따른, SACAE(Self-Attention based Convolutional Autoencoder)와 TCN(Temporal Convolutional Network)을 이용한 다변량 대 다변량 시계열 데이터 예측 모델 생성 방법은 컴퓨팅 장치에 의해 수행되고, 제1 학습 데이터를 이용하여 SACAE(Self-Attention based Convolutional Autoencoder)를 학습하는 단계, 및 제2 학습 데이터를 이용하여 TCN(Temporal Convolutional Network)을 학습하는 단계를 포함하고, 상기 예측 모델은 상기 SACAE와 상기 TCN을 포함하고, 상기 TCN은 상기 SACAE의 인코더의 출력인 잠재 벡터를 입력으로 수신하여 잠재 벡터의 미래 예측값인 예측 잠재 벡터를 출력하고, 상기 SACAE의 디코더는 상기 예측 잠재 벡터를 입력으로 수신하여 재구성하는 것을 특징으로 한다.
- [0009] 또한, 본 발명의 일 실시예에 따른, 시계열 데이터의 예측 방법은 컴퓨팅 장치에 의해 수행되고, 예측 모델에 포함되는 SACAE(Self-Attention based Convolutional Autoencoder)의 인코더를 이용하여 타겟 데이터에 대한 잠재 벡터를 생성하는 단계, 상기 예측 모델에 포함되는 TCN(Temporal Convolutional Network)을 이용하여 상기 잠재 벡터에 대한 미래 예측값인 예측 잠재 벡터를 생성하는 단계, 및 상기 SACAE의 디코더를 이용하여 상기 예측 잠재 벡터를 디코딩하는 단계를 포함한다.

발명의 효과

- [0010] 본 발명의 실시예에 따른 다변량 시계열 데이터 예측 장치 및 방법에 의한 경우, 공정 신호와 같은 다변량 시계열 신호의 미래값을 더욱 정확하게 예측할 수 있는 효과가 있다.

도면의 간단한 설명

- [0011] 본 발명의 상세한 설명에서 인용되는 도면을 보다 충분히 이해하기 위하여 각 도면의 상세한 설명이 제공된다.
- 도 1은 MTS 비선형적인 관계 추출을 설명하기 위한 도면이다.
- 도 2는 CNN의 컨벌루션 커널(convolutional kernel)을 도시한다.
- 도 3은 본 발명에서 제안하는 셀프 어텐션 모듈을 도시한다.
- 도 4는 TCN, 인과 컨벌루션, 및 확장 컨벌루션의 아키텍처를 도시한다.
- 도 5는 본 발명에서 제안하는 모델의 전체적인 구조도를 도시한다.

발명을 실시하기 위한 구체적인 내용

- [0012] 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.
- [0013] 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.
- [0014] 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.
- [0015] 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.

[0016] 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.

[0017] 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.

[0018] 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.

[0019] 본 발명은 다변량 시계열 데이터의 예측을 위한 알고리즘에 관한 것이다. 보다 구체적으로, 다변량 대 다변량 시계열 데이터 예측 모델 생성 방법, 다변량 대 다변량 시계열 데이터 예측 방법, 및 상기 방법들을 수행하는 장치 등이 개시된다. 또한, 상기 방법은 적어도 프로세서(processor) 및/또는 메모리(memory)를 포함하는 컴퓨팅 장치에 의해 수행될 수 있다. 컴퓨팅 장치는 PC(Personal Computer), 태블릿 PC, 서버(Server) 등으로 구현될 수 있다.

[0020] 제조 공정 등에서, 데이터는 상관된 시퀀스(correlated sequences)의 형태로 수집된다. 다변량 대 다변량 시계열(Multivariate to Multivariate Time Series, MMTS) 예측은 제조 공정에서 매우 중요한 요소이다. MMTS 예측(MMTS forecasting)은 미래 시간을 예측하면서 변수와 변수 간의 비선형적인 관계(Inter-relationship)와 각 단변량 시계열 내의 시간적 관계(Intra-relationship)를 모두 통합하여 예측을 수행하는 작업이다. 그러나, 기존의 연구들은 저차원의 데이터(low-dimensional data)에 적합한 딥러닝 모델(deep learning models)을 이용하였다. 이러한 모델들은 MTS(Multivariate Time Series) 데이터에 내재된 고차원의 관계를 모델링하는 데 충분하지 않다. 또한, 이러한 모델들은 멀티플 인풋 변수들(multiple input variables)로부터 단일의 타겟 변수만을 예측하는 데 집중하기 때문에 비생산적임과 동시에 비효율적이다.

[0021] 따라서, 본 발명에서는 두 단계(two phase)로 구성된 MTS 예측 기법 및 장치를 제안한다. 첫째로, 제안하는 모델은 SACAE(Self-Attention based Convolutional Autoencoder)를 통해 UTS(Univariate Time Series, 단변량 시계열) 사이의 비선형 관계(non-linear correlations, 즉 inter-relationship)를 학습하고 원인 분석(cause analysis)을 수행한다. 둘째로, TCN(Temporal Convolutional Network)을 통해 MTS 데이터의 시간적 관계(temporal relationships, 즉 intra-relationships)를 학습하고 멀티플 타겟 출력(multiple target outputs)을 예측한다. 종단간 모델(end-to-end model)로써, 제안하는 기법은 보다 효율적이고 매우 뛰어난 실험적 결과를 도출하였다.

[0022] 또한, 제안하는 기법은 다변량 데이터 예측값을 하나의 예측 변수를 예측하는 것이 아닌, 다 대 다, 다중의 단변량 시계열 데이터를 참고하여 차원을 유지한채 다중의 단변량 시계열을 예측하는 것을 목표로 한다.

[0023] 후술되는 본 발명의 제1 단계(Phase I)의 입력은 MTS 데이터 $\mathbf{X}^i = \{x_0^i, \dots, x_t^i\}$ 이다. 여기서, $\mathbf{X}^i \in \mathbb{R}^{i \times t}$ 는 MTS 데이터 내의 i번째 변수의 타임 포인트들(time points)이다. 제2 단계(Phase II)에서 TCN의 입력은 $\mathbf{X}' = \{x_0', \dots, x_t'\}$ 이다. 여기서, $\mathbf{X}' \in \mathbb{R}^t$ 는 제1 단계(Phase I)의 $\mathbf{X}^i$ 에 대한 압축된 또는 응축된 UTS 표현(condensed UTS representation, 실시예에 따라 잠재 벡터(Latent Vector)로 명명될 수도 있음)이다. MMTS 예측 태스크는 주어진 MTS 입력 데이터로부터 대응하는 $\hat{y}_{t+a}^i$ 를 예측하는 것이다. 그라운드 투르쓰 벨류(ground truth value)는 $\mathbf{X}' = \{x_0', \dots, x_t'\}$ 로 표현된다. 수학적 1은 아래와 같다.

[0024] [수학식 1]

$$[0025] \hat{y}_{t+a}^i = x_{t+a}^i = f(x_0^i, \dots, x_t^i; \theta) + \varepsilon$$

[0026] 수학식 1에서, $f(\cdot)$ 는 모델을, a 는 미래 타임 스텝들의 범위(range of future time steps)를, θ 는 학습가능한 파라미터(learnable parameter)를, ε 는 모델의 예측 에러(prediction error)를 나타낸다.

[0027] 이하에서는, 본 발명을 구현하기 위한 CAE(Convolutional Autoencoder), 셀프 어텐션(self-attention), 및 TCN(Temporal Convolutional Network)에 대해 설명한다.

[0028] **CAE(Convolutional Autoencoder)**

[0029] CAE는 재구성(reconstruction)과 특징 추출(feature extraction, 또는 특질 추출)을 목적으로 하는 신경망(neural network)의 변형 모델이다. CNN(Convolutional Neural Network)은 입력 이미지들로부터 주요 특징들(key features)을 드라마틱하게 생성할 수 있다. 오토인코더(Autoencoder, AE)는 특징 벡터들(feature vectors)을, 고차원의 데이터 공간(high dimensional data space)으로부터 저차원의 데이터 공간(low dimensional data space)으로의 비선형적인 관계(inter-relationship)를 학습할 수 있는 추상적인 특징 벡터들(abstract feature vectors)로 변환할 수 있다.

[0030] Zheng 등(2014)은 분류 성능을 향상시키기 위하여 MTS를 UTS로 분할하고 심층 CNN(deep Convolutional Neural Network)을 이용하여 MTS 데이터에 대한 특징 학습 기법을 제안한 바 있다. 그러나, Zheng 등이 제안한 기법은 UTS 간의 비선형적인 관계를 포함할 수 없다는 한계점이 있다. 이러한 문제점을 극복하기 위하여, CNN과 AE에 영감을 받은 MMTS 예측을 위한 프레임워크가 제안되었다. CNN과 AE 두 모델의 결합, 즉 CAE는 입력의 로컬 맵(local map)의 대표 특징들(representative features)을 추출한다. MTS를 UTS로 분할하는 대신, 특징 추출을 위해 MTS를 공동으로 이용한다. CNN의 가중치 공유 구조 특성(weight sharing structural characteristics)으로 인하여, CAE는 적은 학습 파라미터들과 높은 학습 효율을 갖고 과적합(over-fitting)을 피할 수 있다.

[0031] 제안하는 기법에서, CAE는 MTS 데이터의 멀티플 변수들의 비선형적인 관계를 추출하기 위해 이용된다. AE를 통한 로컬 컨벌루션을 추가하여, CAE는 입력의 로컬 특징들(local features)을 특징 맵(feature map)으로 추출함으로써, 오리지널 입력을 인코딩할 수 있다. 그러면, 컨벌루션 디코더는 추출된 특징 맵을 오리지널 데이터로 디코딩할 수 있다. 로컬 특징 획득 특성으로 인하여, 다양한 특징들을 추출하기 위하여 CAE의 동일한 컨벌루션 레이어들에서 멀티플 필터들이 이용된다. MTS 비선형적인 관계 추출은 도 1에 도시되어 있다.

[0032] CAE의 일반적인 컨벌루션 동작은 수학식 2와 같이 표현된다.

[0033] [수학식 2]

$$[0034] \mathbf{Z}_c = h(\mathbf{W}_c * \mathbf{X}^i + \mathbf{b}_c)$$

[0035] 수학식 2에서, $\mathbf{W}_c$ 와 $\mathbf{b}_c$ 는 가중치 파라미터들(weight parameters)이다. $h(\cdot)$ 는 활성화 함수(activation function)이고, $*$ 는 컨벌루션 레이어와 픽셀들(pixels)의 도트 프덕트(dot product)를 나타낸다. c 번째 필터는 입력 행렬(input matrix) $\mathbf{X}^i$ 를 스캔하고 $\mathbf{Z}_c$, 컨벌루션 연산의 출력을 생성한다.

[0036] 컨벌루션 연산 이후에, PReLU(Parametric Rectified Linear Unit)의 활성화 함수가 이용된다. PReLU는 음수값을 위한 기울기(slope)를 가진 전통적인 rectified unit을 생성하는 활성화 함수이다. PReLU는 수학식 3과 같이 표현된다.

[0037] [수학식 3]

$$[0038] h(\mathbf{T}_c) = \begin{cases} \mathbf{T}_c, & \text{if } \mathbf{T}_c > 0 \\ p_c \mathbf{T}_c, & \text{if } \mathbf{T}_c \leq 0 \end{cases}$$

[0039] $\mathbf{T}_c = \mathbf{W}_c * \mathbf{X}^i + \mathbf{b}_c$ 는 활성화 함수 $h(\cdot)$ 에 대한 c 번째 채널 입력이고, p_c 는 활성화 함수의 음수 부분(negative part)의 제어를 위한 계수(coefficient)이다. p_c 가 0이 되면, 활성화 함수는 ReLU가 된다. p_c 가 학습가능한 파라미터(learnable parameter)일 때, 수학식 3은 PReLU이다.

[0040] 다만, 본 발명이 사용하는 활성화 함수에 제한되는 것은 아니며, ReLU 또는 다른 활성화 함수가 사용될 수도 있

다.

[0041] Self-attention(셀프 어텐션)

[0042] CNN과 CNN의 로컬 컨벌루션 커널은 컨벌루션 연산이 커널의 오직 지역적 정보(local information)만을 획득(capture)하도록 제한한다. 도 2는 CNN의 컨벌루션 커널(convolutional kernel)을 도시한다. 커널 사이즈가 1×3 이라면, 결과는 수용 범위(receptive field, 수용장이라고 명명될 수도 있음)의 3개의 값에만 의존하여 결정된다. 더 많은 정보를 반영하기 위해, 컨벌루션들은 수차례 적층되거나, 풀링 레이어(pooling layer)를 통해 데이터의 사이즈가 감소될 수 있다. 그러나, 수용 범위를 늘리는 것은 비지역적 의존성 문제(non-local dependency problem)를 해결하기 위한 효율적인 기법은 아니다.

[0043] Wang 등(2018)은 셀프 어텐션 모듈 타입인 비지역적 신경망(non-local neural network)를 제안한 바 있다. 셀프 어텐션 연산을 구현함으로써, 본 발명에서 제안하는 기법은 임의의 두 지점 간의 인터랙션만을 통하여 비지역적 의존성을 캡처할 수 있다. 또한, 셀프 어텐션 모듈로부터의 어텐션 맵(attention map)을 분석하는 것은 MMTS 예측 이후의 인과 분석을 가능케 한다. 본 발명에서 제안하는 기법의 셀프 어텐션 모듈은 도 3에 도시되어 있다.

[0044] 셀프 어텐션 모듈의 출력은 수학적 4와 같이 표현된다.

[0045] [수학적 4]

$$[0046] \mathbf{X}' = \mathbf{Z}_c + \mathbf{O}_c$$

[0047] 수학적 4를 참조하면, 셀프 어텐션의 출력 특징 맵 $\mathbf{O}_c \in \mathbb{R}^{1 \times W \times C}$ 은 오리지널 입력 $\mathbf{Z}_c \in \mathbb{R}^{1 \times W \times C}$ 에 더해진다.

[0048] 1, W, 및 C는 컨벌루션 연산의 결과와 출력 특징 맵의 높이(height), 폭(width), 및 채널을 나타낸다. 어텐션 특징은 CAE의 인코더와 디코더의 마지막 레이어와 첫번째 레이어에서 셀프 어텐션이 삽입되도록 한다. 다만, 본 발명이 셀프 어텐션이 삽입되는 위치 및 삽입되는 개수에 제한되는 것은 아니다. 실시예에 따라, 셀프 어텐션은 CAE의 인코더 내에 복수회로 삽입되거나, 및/또는 CAE의 디코더 내에 복수회로 삽입될 수도 있다. 셀프 어텐션 모듈은 수학적 5와 같이 계산된다.

[0049] [수학적 5]

$$[0050] \Phi(z_i, z_j) = e^{k(z_i)^T \cdot q(z_j)}$$

[0051] 수학적 5에서, z_i 와 z_j 는 특정 컨벌루션 레이어와 $k(\cdot)$ 의 특징 맵들이고, $q(\cdot)$ 는 1×1 컨벌루션들이다. T는 변수를 전치한다(transpose). 특징 맵들에 1×1 컨벌루션들을 적용함으로써, 채널의 개수와 계산 비용을 감소시킬 수 있다. 어텐션 가중치들(attention weights)은 수학적 5의 출력 스코어를 정규화함으로써(normalizing) 계산될 수 있다. 어텐션 가중치들은 수학적 6에 의해 계산된다.

[0052] [수학적 6]

$$[0053] \psi(i, j) = \frac{\Phi(z_i, z_j)}{\sum_{k=1}^{1 \times W} \Phi(z_i, z_k)}$$

[0054] 다음으로, 출력 특징 맵 $\mathbf{O}_c$ 에 적용되는 어텐션을 생성하기 위하여 어텐션 가중치가 입력 특징 맵을 이용하여 계산된다. 출력 특징 맵을 위한 수학적 7과 같이 표현된다.

[0055] [수학적 7]

$$[0056] \mathbf{O}_c = \sum_{j=1}^w \psi(z_i, z_j) \cdot v(z_j)$$

[0057] 수학적 7에서, $v(\cdot)$ 는 1×1 컨벌루션이다.

[0058] 수학적 5를 살펴보면, 컨벌루션 레이어의 특징 맵들 사이의 관계에 기초하여 계산되기 때문에, 셀프 어텐션 모듈은 지역적 의존성 이슈를 완화시킬 수 있다.

[0059] TCN(Temporal Convolutional Network)

[0060] TCN은 시퀀셜 데이터(sequential data)를 위한 컨벌루션 아키텍처의 변형예이다. TCN은 인과 컨벌루션들(causal convolutions)과 확장 컨벌루션들(dilated convolutions)을 포함한다.

[0061] MMTS 예측을 위해 TCN은 응축된 UTS 표현(condensed UTS representation)으로부터 오리지널 MTS 입력을 재구성한다. 그러나, LSTMs과 GRUs는 사라지는(vanishing)/폭증하는(exploding) 기울기 문제(gradient problems)를 갖는다. 이는 응축된 UTS 표현의 정확한 예측에 부적절하다.

[0062] TCN은 이러한 모델들에 비하여 몇가지 장점을 갖는다. TCN의 중요한 장점 중 하나는 기울기를 안정화시키는 능력이다. TCN은 사라지는/증가하는 기울기 문제를 회피할 수 있다.

[0063] TCN의 모델 구조를 설명하기 전에, 응축된 UTS 표현의 시퀀스 모델링을 정의한다. 응축된 UTS 표현 $\mathbf{X}' = \{x'_0, \dots, x'_t\}$ 은 예상되는 출력 $x'_{t+1}, \dots, x'_{t+a}$ 를 예측하기 위해 훈련된다. a 는 미래 타임 스텝들의 범위이다. 시퀀스 예측의 핵심은 미래 입력 시퀀스 $x'_{t+1}, \dots, x'_{t+a}$ 가 아닌 주어진 과거 입력 시퀀스 $x'_0, \dots, x'_t$ 에만 의존하여 출력 x'_{t+a} 를 예측한다는 것이다. 일반적인 함수는 수학적 식 8과 같다.

[0064] [수학적 식 8]

[0065]
$$x'_{t+a} = r(x'_0, \dots, x'_t)$$

[0066] 수학적 식 8에서, $r(\cdot)$ 은 TCN 모델을 나타낸다.

[0067] 인과 컨벌루션들(causal convolutions)은 시계열 데이터의 본성, 즉 $t+a$ 시점에서의 미래의 출력은 미래의 데이터 포인트가 아닌 오직 현재와 과거에만 의존한다는 인과 관계(causality)를 만족하여야 한다는 것을 만족한다. TCN은 시점 t 와 이전 레이어 내의 이전 시점으로부터의 요소들만과의 컨벌루션을 달성하기 위해 커널 사이즈 -1을 갖는 제로 패딩(zero-padding)을 갖는 1-D 컨벌루션을 이용한다. 그러나, 1-D 컨벌루션들과 인과 컨벌루션들은 동떨어진 타임 이벤트들(distant time events)을 캡처하기 위해 많은 레이어들과 큰 필터들을 필요로 한다. 많은 레이어들과 큰 필터들을 요구하는 비효율성을 해소하기 위해, TCN은 확장 컨벌루션들(dilated convolutions)을 이용한다. Wavenet(Oord et al., 2016)에 의해 진보되어, 확장 컨벌루션들은 저 계산 비용을 갖는 크기의 차수(order of magnitude)에 의해 수용 범위(receptive field)를 증가시킨다. 확장 필터(dilation filter)의 사이즈는 수학적 식 9와 같다.

[0068] [수학적 식 9]

[0069]
$$Dilation\ filter = (p - 1)\Sigma_l d_l + 1$$

[0070] 수학적 식 9에서, 레이어 l 의 확장 사이즈는 d_l 이고, 필터 사이즈는 p 이고, 확장률은 d 이다. 컨벌루션 커널들 내에 0(zeros)을 삽입함으로써, $d-1$ 스킵된 입력이 컨벌루션에 적용된다. TCN, 인과 컨벌루션, 및 확장 컨벌루션의 아키텍처는 도 4에 도시되어 있다. 복수의 확장 컨벌루션 레이어들을 적층(stack)함으로써, TCN은 오리지널 컨벌루션 동작에 비해 더 적은 레이어들을 갖는 큰 수용 범위를 달성할 수 있다.

[0071] 이하에서는, 본 발명에서 제안하는 기법을 설명한다.

[0072] 본 발명에 의한 방법은 크게 SACAE를 포함하는 제1 단계(Phases 1)와 TCN을 포함하는 제2 단계(Phase 2)로 구성된다. 두 단계를 이용한 예측은 오리지널 입력 데이터의 복잡한 대표 특징들(representative features)을 모델이 순차적으로 학습하도록 한다.

[0073] 제1 단계(Phase 1)는 SACAE를 통해 MTS 데이터의 비선형적인 관계(inter-relationships)를 추출한다. 그런 다음, MTS 데이터의 응축된 UTS(univariate time series) 표현(representation)을 생성한다. CAE 과정 동안, 적어도 CAE의 인코더의 마지막 컨벌루션 레이어 및 CAE의 디코더의 첫번째 컨벌루션 레이어에 삽입된 셀프 어텐션들로 인하여 지역 정보(local information)에 의존하는 문제점을 완화할 수 있다. 학습된 SACAE 인코더는 MTS 데이터의 응축된 UTS 표현(representation)을 제2 단계(Phases 2)로 전달한다. 그러면, 데이터의 시간적 관계(intra-relationship)에 대해 학습된 TCN은 단변량 표현 시계열(univariate representation time series)의 미래 타임 스텝들을 예측한다. 예측된 타임 스텝은 MMTS 동작과 인과 분석을 위해 SACAE의 학습된 디코더에 의해 디코딩된다. 본 발명에서 제안하는 기법의 구조는 도 5에 도시되어 있다.

[0074] 이하에서, 제1 단계와 제2 단계에 대해 보다 구체적으로 설명한다.

[0075] 제1 단계(Phases I)

[0076] 제1 단계는 학습 데이터를 이용하여 SACAE를 학습하는 단계를 의미할 수 있다. 즉, SACAE는 입력 데이터와 출력 데이터 간의 재구성 오류가 작아지는 방향으로 학습될 수 있다. 학습 데이터는 유무선 통신망을 통해 컴퓨팅 장치에 의해 수신된 후 컴퓨팅 장치에 포함된 저장부에 저장될 수 있다. 실시예에 따라, 학습 데이터는 USB 저장 장치와 같은 저장 장치로부터 소정의 입력 인터페이스를 통해 수신될 수도 있다.

[0077] 본 발명에서 제안하는 기법은 MTS 데이터 내에 존재하는 비선형 코릴레이션(non-linear correlation)을 나타내는 비선형적인 관계(inter-relationship)를 포착하기 위해서 SACAE를 이용한다. SACAE 인코더는 멀티플 1-D 컨벌루션 블럭들(multiple 1-D convolutional blocks)과 셀프 어텐션 모듈들(self-attention modules)로 구성될 수 있다. 간단한 벡터 아웃풋(vector output)을 출력하기 위하여 플래튼 레이어(flatten layer)가 마지막 동작에 추가될 수 있다.

[0078] MTS 데이터는, 커널 사이즈(kernel size)는 3이고, 스트라이드 사이즈(stride size)는 1이고, 패딩 사이즈(padding size)는 1인 1-D 컨벌루션 동작을 갖는 채널들로써의 SACAE에 입력된다. 이로 인하여, SACAE는 각 컨벌루션 블럭 동작들 이후의 다음 출력들의 전체적인 길이를 보존할 수 있다. 입력과 동일한 길이를 갖는 출력은, 제2 단계에서 TCN이 다음 타임 스텝을 예측할 수 있도록 한다.

[0079] 데이터가 SACAE 인코더의 각 컨벌루션 블럭으로 전달됨에 따라, MTS 입력 데이터의 풍부한 특징들(abundant features)을 추출하기 위한 채널의 개수는 기하급수적으로 감소한다. SACAE 인코더의 각 컨벌루션 블럭은 입력 채널의 개수를 감소시킨다. 그러면, 생성된 채널들은 입력 MTS 데이터의 개별 변수들(individual variables) 사이의 풍부한 정보를 추출한다. 제조 공정에서 동일한 기계(machinery)로부터 생성된 데이터가 일 예이다. CAE가 로컬 필드에서 MTS 데이터의 비선형 코릴레이션들을 획득하는 반면에, 비지역 의존 이슈(non-local dependency issue)는 글로벌 의존을 추출하는 것을 어렵게 한다. 이러한 문제점을 완화하기 위하여, 셀프 어텐션이 CAE의 (인코더의) 마지막 레이어와 (디코더의) 첫번째 레이어에 추가될 수 있다. 따라서, SACAE는 특징 맵(feature map)에서의 요소들의 중요성을 계산하고 글로벌 의존성을 포착할 수 있다. 얇은 컨벌루션 레이어들(shallow convolutional layers)은 입력의 로컬 특징들을 캡처하는 경향이 있기 때문에, 글로벌 정보를 캡처하기 위하여 컨벌루션 블럭들의 더 깊은 레이어들에 셀프 어텐션이 포함될 수도 있다.

[0080] 제2 단계(Phases II)

[0081] 제2 단계는 TCN을 학습시키는 단계를 의미할 수 있다. TCN은 잠재 벡터의 미래값을 예측하여 예측된 잠재 벡터를 출력한다. TCN의 학습 과정에서 이용되는 제2 학습 데이터는 학습된 SACAE를 통해 도출된 잠재 벡터일 수 있다. 즉, 현재 시점까지의 공정 데이터에 기초한 잠재 벡터와 미래 시점에서의 잠재 벡터는 SACAE를 통해서 도출될 수 있다. SACAE는 수신된 또는 저장된 학습 데이터(SACAE를 학습하는 과정에서 이용되는 학습 데이터를 의미하거나, 별도의 학습 데이터를 의미할 수 있음)를 이용하여 복수의 시점에서의 잠재 벡터를 생성하고, 생성된 잠재 벡터는 TCN의 학습에 이용되는 제2 학습 데이터가 될 수 있다. TCN은 입력 데이터와 예측 데이터 간의 오차가 작아지는 방향으로 학습될 수 있다.

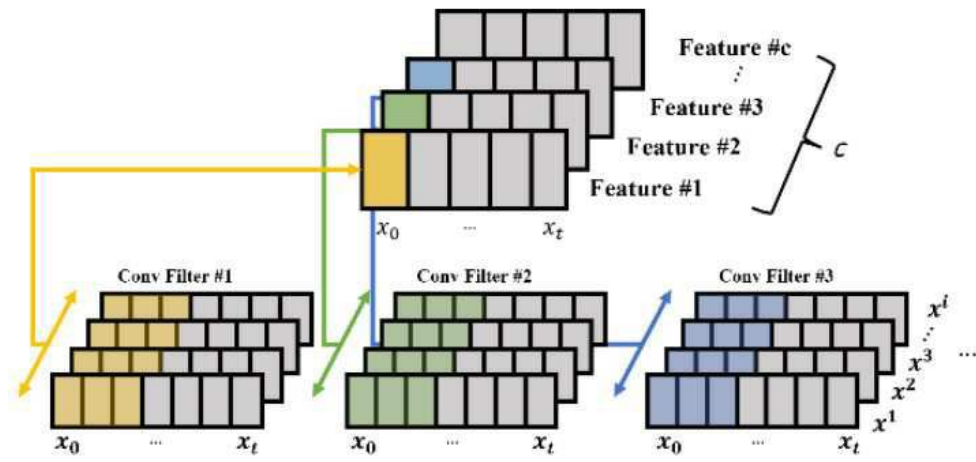
[0082] TCN은 인과 컨벌루션들(causal convolutions)과 확장 컨벌루션들(dilated convolutions)로 구성된다. 확장 컨벌루션들은 연산 효율성(computational efficiency)을 증대시키고, 인과 컨벌루션들은 연관된 히스토리 정보(historical information)를 추출하고 미래의 출력을 예측한다. 제안하는 기법에서 몇몇의 수정이 가미된 베이스라인 TCN(baseline TCN)이 이용된다. TCN의 역할은 오리지널 MTS 데이터의 인코딩된 UTS 표현의 다음 타임 스텝들을 예측하는 것이다. 모듈은 멀티플 템퍼럴 컨벌루션 레이어들(multiple temporal convolutional layers)에 기초하여 인코딩된 데이터의 다음 타임 스텝들을 예측한다. TCN은 복수의 다수의 은닉 채널들을 갖는 템퍼럴 블럭 레이어들로 구성된다. TCN의 구조는 도 3에 도시되어 있다. 확장률(dilation rate) d 는 2^i 의 비율로 증가될 수 있다. 따라서, 레이어들의 개수에 따라, 필터는 전체 입력 시퀀스를 커버할 수 있다.

[0083] SACAE 디코더는 UTS 표현 데이터의 예측된 타임 스텝들을 오리지널 입력 데이터로 재구성한다. 디코더의 셀프 어텐션 블럭의 분석을 통해, 입력 데이터에 대한 출력 데이터의 효과에 대한 인과 분석이 가능하다. SACAE 디코더는 SACAE 인코더의 구조와 유사하다. 멀티플 컨벌루션 블럭들과 셀프 어텐션 레이어들이 임베딩된다. 인코더와 디코더의 전체 구조가 유사하여, SACAE 디코더는 전치된 1-D 컨벌루션들(transposed 1-D convolutions)을 이용할 수 있다.

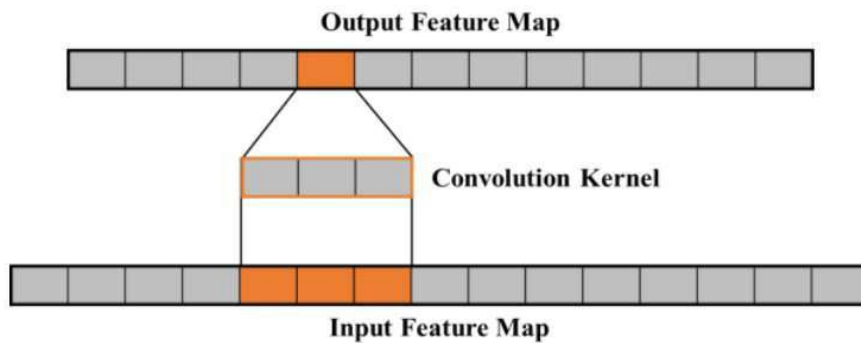
- [0084] 학습이 완료된 모델, 즉 예측 모델은 타겟 데이터(다변량 시계열 데이터(들)을 의미할 수 있음)을 입력받아 미래의 예측값들을 도출할 수 있다. 구체적으로, 예측 모델에 포함된 SACAE의 인코더는 타겟 데이터를 수신하여 잠재 벡터(들)을 생성한다. 생성된 잠재 벡터(들)은 예측 모델에 포함된 TCN에 입력되고, TCN은 잠재 벡터(들)에 대한 예측값(들), 즉 예측된 잠재 벡터(들)을 출력한다. 예측된 잠재 벡터(들)은 다시 SACAE의 디코더에 입력되고 최종적인 예측값(들)이 출력될 수 있다. 예측 모델을 이용한 예측 동작은 앞선 기재와 중복되기 때문에 구체적인 설명은 생략하기로 한다.
- [0085] 최종적인 예측값(들)을 이용하면 설비의 이상 유무를 판단하거나 설비의 수명을 예상할 수 있다.
- [0086] 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, 컨트롤러, ALU(Arithmetic Logic Unit), 디지털 신호 프로세서(Digital Signal Processor), 마이크로컴퓨터, 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 컨트롤러를 포함할 수 있다. 또한, 병렬 프로세서(Parallel Processor)와 같은, 다른 처리 구성(Processing Configuration)도 가능하다.
- [0087] 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.
- [0088] 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.
- [0089] 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

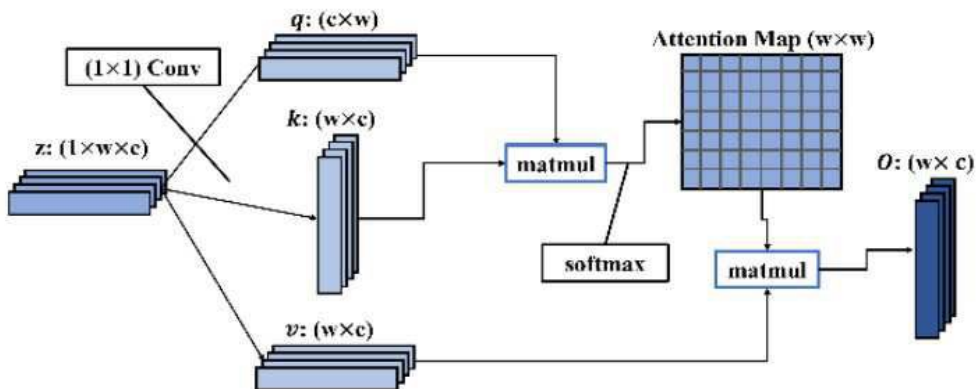
도면1



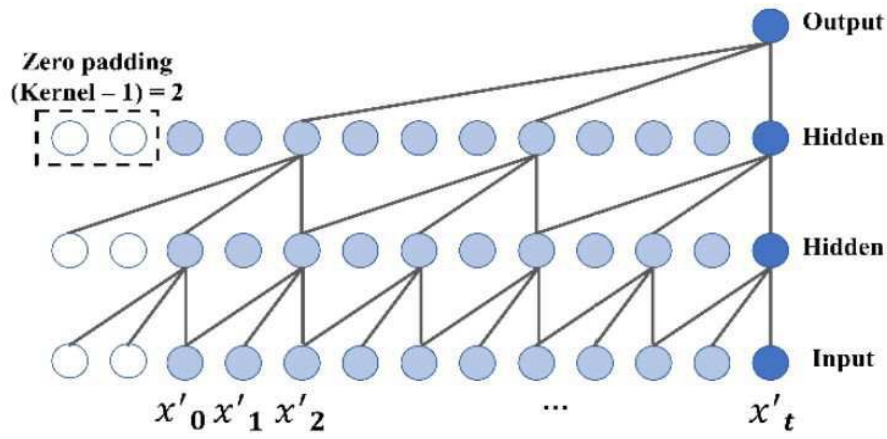
도면2



도면3



도면4



도면5

