2024년 3월 8일 DMQA 연구실 오픈 세미나

Diffusion Models for Time Series

Data Mining & Quality Analytics Lab.

추창욱

발표자 소개

- **추창욱 (Changwook Chu)**
    - 고려대학교 산업경영공학과 대학원 재학
    - Data Mining & Quality Analytics Lab. (김성범 교수님)
    - M.S. Student (2023.09 ~ Present)

- **Research Interest**
    - Diffusion Models
    - Time Series Imputation & Forecasting

- **Contact**
    - chaliechu117@korea.ac.kr

Contents

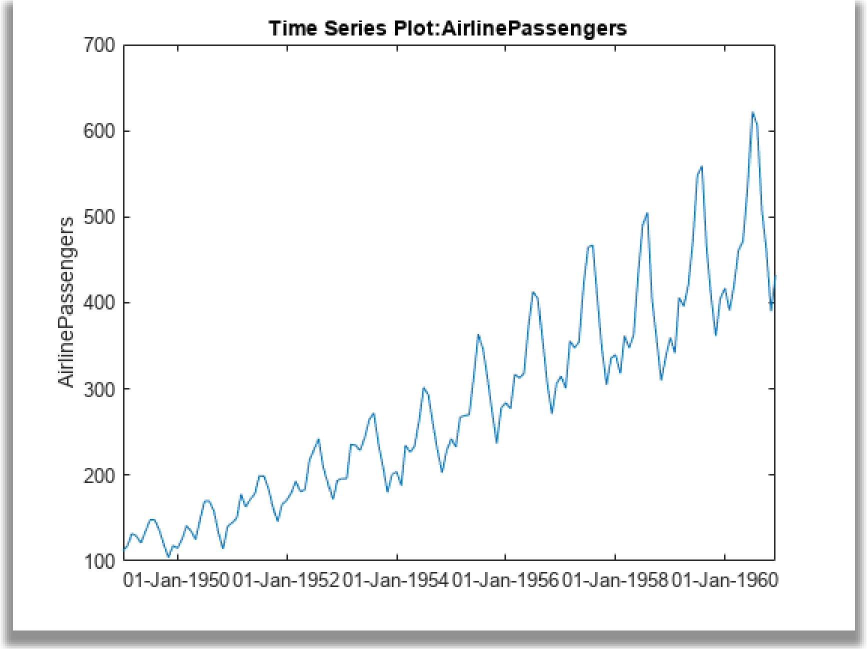
Introduction

Introduction

What is Time Series Data?

❖ 시계열 데이터 (Time Series Data)

- 시간 순서에 따라 관측된 데이터
- 주식, 기상 정보, 여러 공정 등 다양한 분야에서 사용
- 자동화 및 센서 기술의 발전, 인터넷의 보편화 등 → 대규모 시계열 데이터의 증가



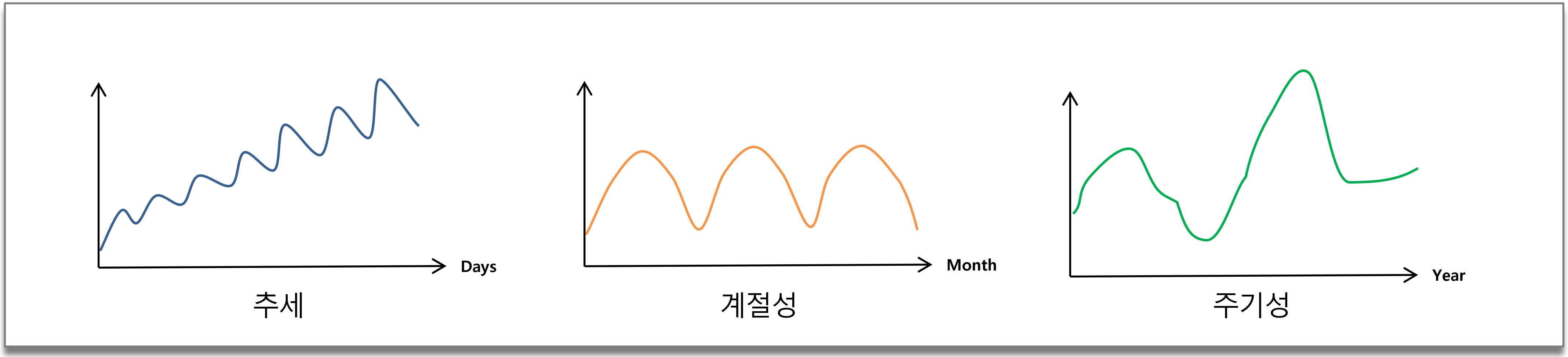
https://www.mathworks.com/

Introduction

What is Time Series Data?

❖ 시계열 데이터 패턴

- 추세(Trend) : 장기적인 추이 또는 증가 또는 감소하는 방향성을 나타냄
- 계절성(Seasonality) : 고정된 빈도로 반복되는 주기적인 패턴을 나타냄
- 주기성(Cycle) : 고정된 빈도가 아닌 형태로 증가나 감소하는 모습을 나타냄



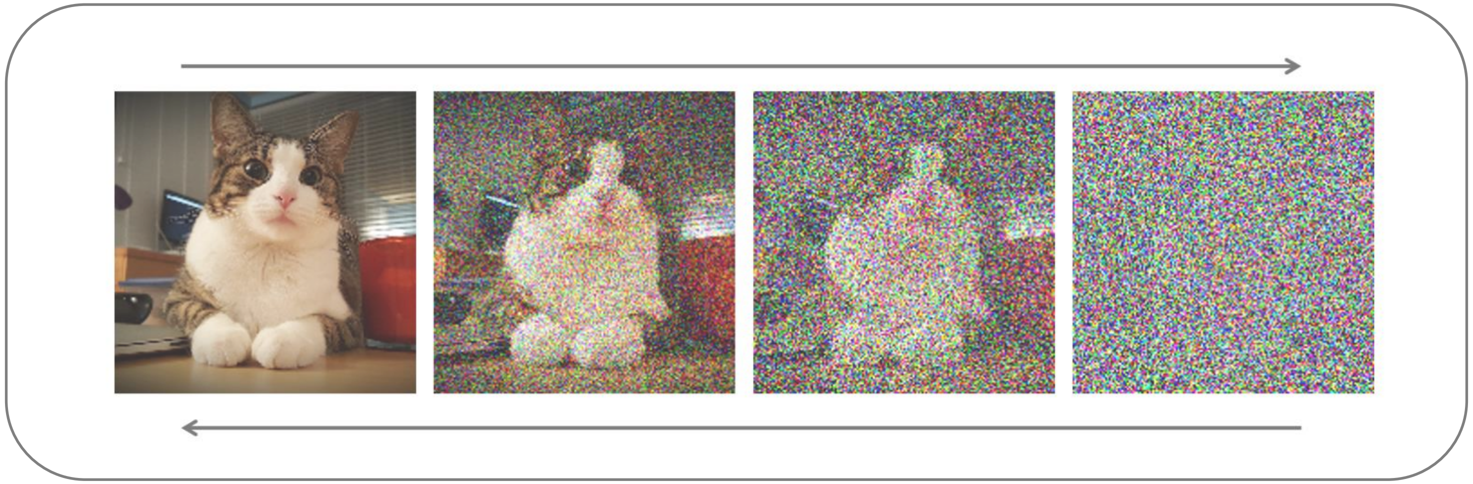
Introduction

What is Time Series Data?

❖ 시계열 데이터 연구의 여러 분야

- Forecasting : 미래의 값이나 발생할 사건을 예측
- Missing Data Imputation : 결측치 대체
- Anomaly Detection : 데이터의 비정상적인 패턴이나 값들을 식별
- Classification : 데이터가 속하는 카테고리를 분류

Let's use Diffusion Models!



https://developer-blogs.nvidia.com/wp-content/uploads/2022/04/Generation-with-Diffusion-Models.png

Introduction

What is Diffusion Model?

❖ Diffusion Probabilistic Model

- Deep Generative Model
- Forward Process : 데이터로부터 노이즈를 조금씩 더해 완전한 노이즈로 만드는 과정
- Reverse Process : 노이즈로부터 조금씩 노이즈를 제거해가며 데이터를 만들어내는 과정

$$q(x^n|x^{n-1}) = N(x^n; \sqrt{1-\beta_n} x^{n-1}, \beta_n I)$$

Forward Process

x^0 x^1 … x^{n-1} x^N

Reverse Process

$$p_\theta(x^{n-1}|x^n) = N(x^{n-1}; \mu_\theta(x^n, n), \Sigma_\theta(x^n, n) I)$$

https://developer-blogs.nvidia.com/wp-content/uploads/2022/04/Generation-with-Diffusion-Models.png

Introduction

❖ Diffusion Probabilistic Model

- Deep Generative Model
- Forward Process : 데이터로부터 노이즈를 조금씩 더해 완전한 노이즈로 만드는 과정
- Reverse Process : 노이즈로부터 조금씩 제거해가며 데이터를 만들어내는 과정

$$q(x^n|x^{n-1}) = N(x^n; \sqrt{1-\beta_n}x^{n-1}, \beta_n I)$$

$$p_\theta(x^{n-1}|x^n) = N(x^{n-1}; \mu_\theta(x^n, n), \Sigma_\theta(x^n, n)I)$$

Why Diffusion Model?

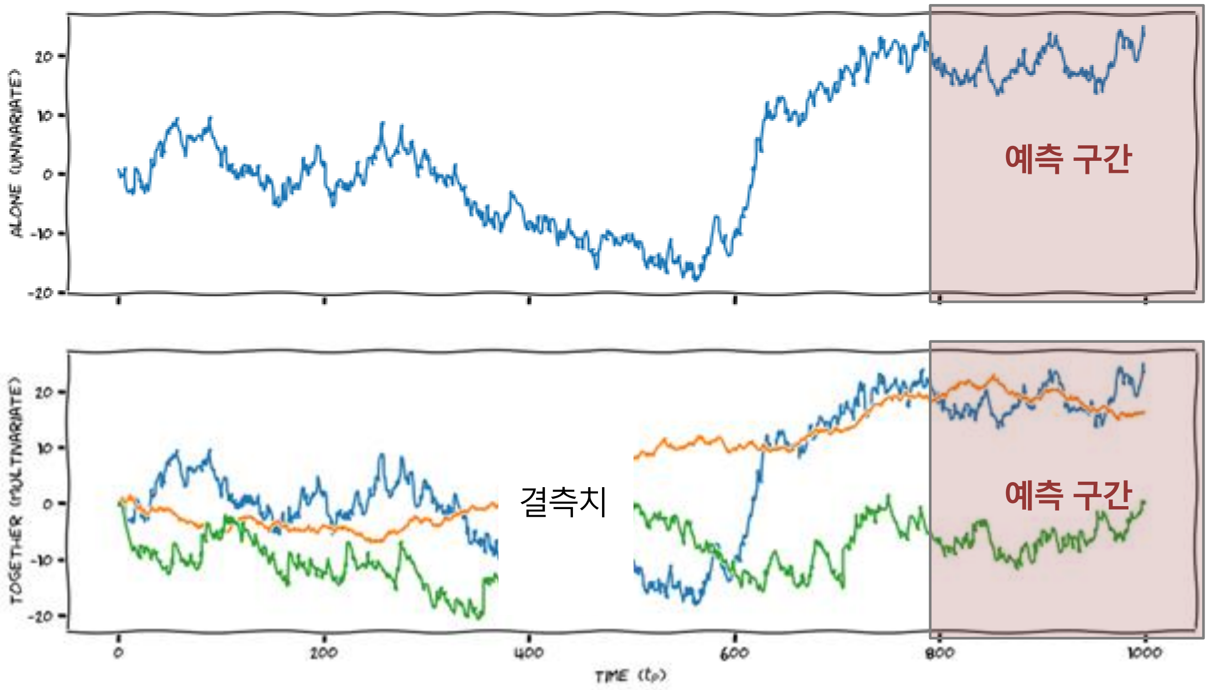
https://developer-blogs.nvidia.com/wp-content/uploads/2022/04/Generation-with-Diffusion-Models.png

Why Diffusion Models?

Time Series Forecasting

❖ 시계열 예측 (Time Series Forecasting)

- 관측된 시계열 데이터를 분석하여 미래를 예측하는 문제



❖ 시계열 예측 Challenges

- 적합한 확률 분포 모델 필요
- 숨겨진 여러 형태의 패턴을 찾아 학습
- 다변량 시계열 데이터 다루기
- 노이즈 또는 결측치 처리

https://miro.medium.com/v2/resize:fit:1204/1*7vvdQ1BznK8Oq2dGdKtQfA.png

Why Diffusion Models?

Univariate Time Series Forecast

❖ 단변량 시계열 (Univariate Time Series) 예측

- 단일 변수의 변화를 모델링하고 예측하는 작업
- 지수평활 모형(exponential smoothing models), ARIMA 모형(autoregressive integrated moving average models) 등

❖ 단변량 시계열 예측의 문제점

- 각 시계열이 개별적으로 학습
- 계절성을 반영하기 위해 Hand-tuned features 필요

Why Diffusion Models?

Deep Learning based Time Series Forecast

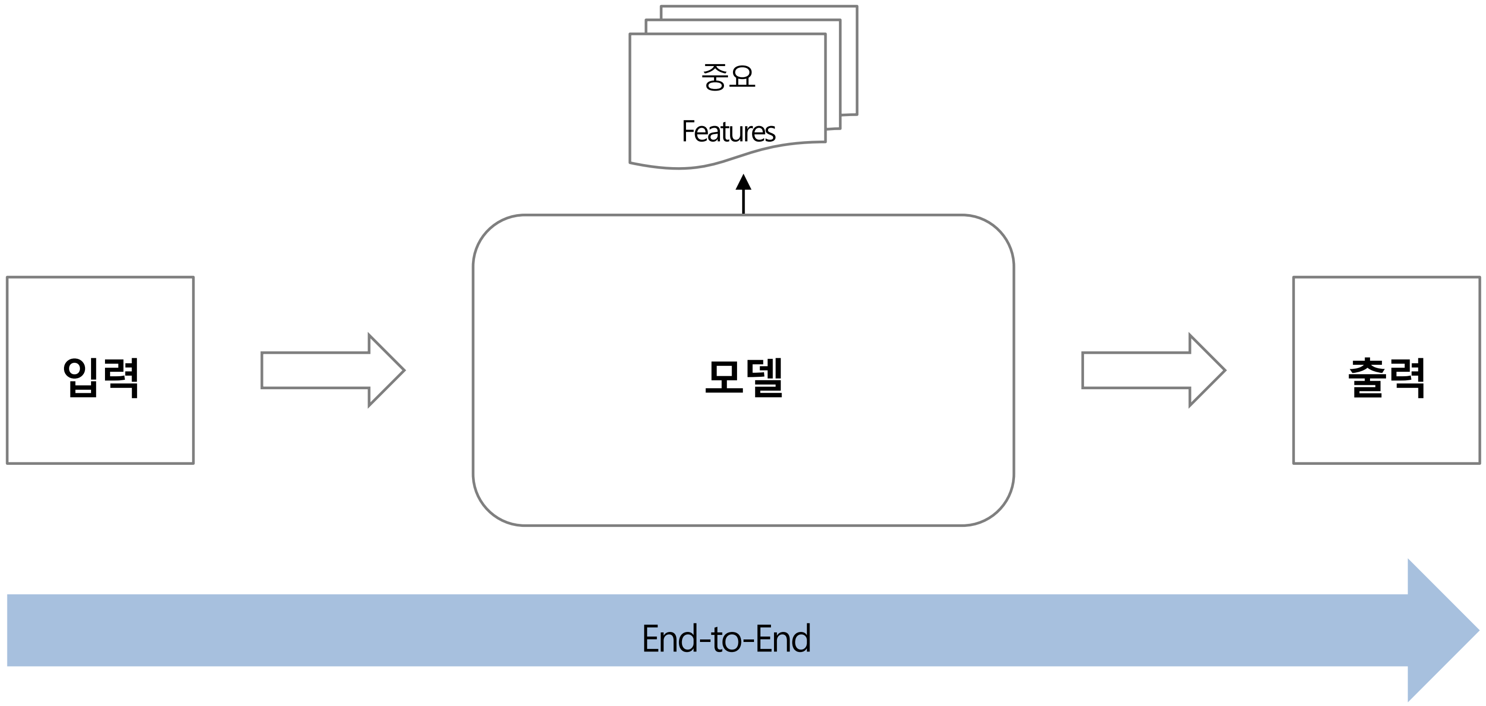
❖ 단변량 시계열 예측의 문제점

✓ 각 시계열이 개별적으로 학습

✓ 계절성을 반영하기 위해 hand-tuned features 필요

❖ 딥러닝 모델의 장점

- 글로벌 모델의 end-to-end 훈련
  - ✓ 데이터 입력부터 출력까지 모든 과정을 한번에 최적화
- 자동 features 추출 능력



Why Diffusion Models?

Deep Learning based Time Series Forecast

❖ 이상적인 딥러닝 기반 시계열 예측 모델

- 대부분 개별 시계열은 서로 통계적 의존성을 가짐

귀납적 편향 통합
Incorporate Inductive bias
학습 데이터로부터 얻은 특정한 가정을 활용하여 예측

+

다변량
Multivariate
여러 변수 간의 상호작용과 의존성 고려

+

확률적 방법
Probabilistic methods
정확한 값과 각 예측에 대한 확률 분포 제공

= **전체 예측 분포를 고려하는 모델**

Why Diffusion Models?

Autoregressive, Generative based normalizing Flow models

귀납적 편향 통합
Incorporate Inductive bias
학습 데이터로부터 얻은
특정한 가정을 활용하여 예측

다변량
Multivariate
여러 변수 간의 상호작용과
의존성 고려

확률적 방법
probabilistic methods
정확한 값과 각 예측에 대한
확률 분포 제공

전체 예측 분포를 고려하는 모델

❖ 대표적인 모델들

- 다변량 시계열 데이터에 대해 더 Flexible한 모델 학습 가능

GAN: 생성자는 실제 데이터와 유사한 데이터를 생성, 판별자는 실제 데이터와 생성된 데이터를 구분하게끔 학습 (경쟁 구도)

시계열 GAN 모델: TimeGAN, C-RNN-GAN, SeqGAN

VAE: 데이터를 잠재공간의 분포로 인코딩하고, 분포로부터 샘플링하여 새로운 데이터 생성

시계열 VAE 예시: VRNN, DeepAR …

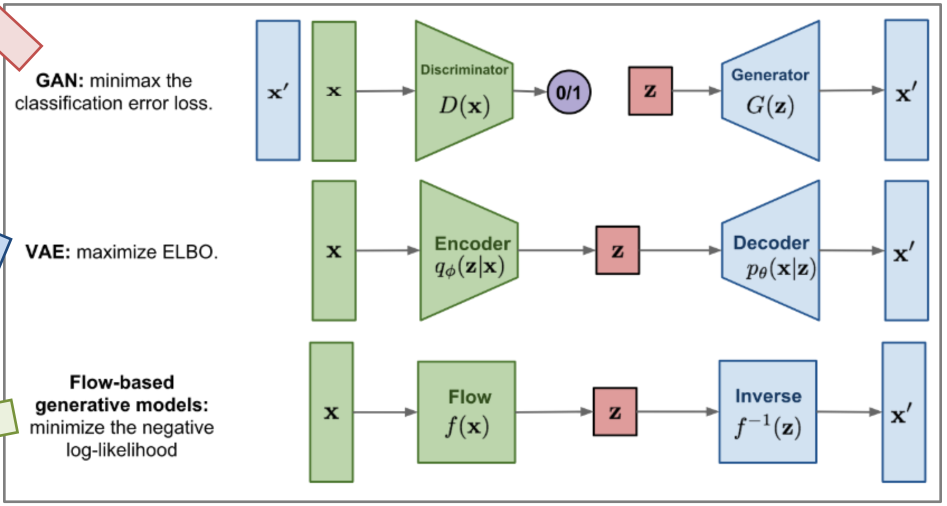
Flow-based: 복잡한 데이터 분포를 간단한 분포로부터 일련의 가역적인 변환을 통해 학습

시계열 Flow-based 예시: NVAE, Glow …

Generative models

Encoder 혹은 다른 **확률적 방법**을 통해

잠재 벡터(z)를 만들고 생성하는 모델링 기법



https://lilianweng.github.io/posts/2018-10-13-flow-models/

$q(x^n|x^{n-1}) = N(x^n; \sqrt{1-\beta_n}x^{n-1}, \beta_n I)$

Forward Process

x^0 x^1 ... x^{n-1} x^N

Reverse Process

$p_\theta(x^{n-1}|x^n) = N(x^{n-1}; \mu_\theta(x^n, n), \Sigma_\theta(x^n, n)I)$

Why Diffusion Models?

Autoregressive, Generative based normalizing Flow models

❖ Diffusion Models의 강점

GAN: 생성자는 실제 데이터와 유사한 데이터를 생성, 판별자는 실제 데이터와 생성된 데이터를 구분하게끔 학습 (경쟁 구도)

시계열 GAN 모델: TimeGAN, C-RNN-GAN, SeqGAN

GAN vs Diffusion

1. 학습의 안정성
2. 높은 품질의 생성물

VAE: 데이터를 잠재공간의 분포로 인코딩하고, 분포로부터 샘플링하여 새로운 데이터 생성

시계열 VAE 예시: VRNN, DeepAR ...

VAE vs Diffusion

1. 생성물의 품질
2. 복잡한 데이터 분포 학습

Flow-based: 복잡한 데이터 분포를 간단한 분포로부터 일련의 가역적인 변환을 통해 학습

시계열 Flow-based 예시: NVAE, Glow ...

Flow-based vs Diffusion

1. 유연성

https://lilianweng.github.io/posts/2018-10-13-flow-models/

Why Diffusion Models?

Autoregressive, Generative based normalizing Flow models

$$q(x^n|x^{n-1}) = N(x^n; \sqrt{1-\beta_n}x^{n-1}, \beta_n I)$$

Forward Process

Reverse Process

$$p_\theta(x^{n-1}|x^n) = N(x^{n-1}; \mu_\theta(x^n, n), \Sigma_\theta(x^n, n)I)$$

❖ Diffusion Models의 강점

GAN: 생성자는 실제 데이터와 유사한 데이터를 생성, 판별자는 실제 데이터와 생성된 데이터를 구분하게끔 학습 (경쟁 구도)

시계열 GAN 모델: TimeGAN, C-RNN-GAN, SeqGAN

→ **GAN vs Diffusion**

1. 학습의 안정성
2. 높은 품질의 생성물

VAE: 데이터를 잠재공간의 분포로 인코딩하고 분포로부터 샘플링하여 새로운 데이터 생성

시계열 VAE 예시: VRNN, DeepAR …

→ **VAE vs Diffusion**

1. 생성물의 품질
2. 복잡한 데이터 분포 학습

Flow-based: 복잡한 데이터 분포를 간단한 분포로부터 일련의 가역적인 변환을 통해 학습

시계열 Flow-based 예시: NVAE, Glow …

→ **Flow-based vs Diffusion**

1. 유연성

강력한 Diffusion Model을 이용해보자!

https://lilianweng.github.io/posts/2018-10-13-flow-models/

Diffusion models for Time Series

선행연구

A versatile Diffusion Model for Audio Synthesis

❖ DiffWave

- 2021년에 제안된 Diffusion Model을 이용한 waveform generation (ICLR, 24년 03월 기준 865회 인용)

DIFFWAVE: A VERSATILE DIFFUSION MODEL FOR AUDIO SYNTHESIS

Zhifeng Kong *
Computer Science and Engineering, UCSD
z4kong@eng.ucsd.edu

Wei Ping
NVIDIA
wping@nvidia.com

Jiaji Huang, Kexin Zhao
Baidu Research
{huangjiaji,kexinzhao}@baidu.com

Bryan Catanzaro
NVIDIA
bcatanzaro@nvidia.com

ABSTRACT

In this work, we propose DiffWave, a versatile diffusion probabilistic model for conditional and unconditional waveform generation. The model is non-autoregressive, and converts the white noise signal into structured waveform through a Markov chain with a constant number of steps at synthesis. It is efficiently trained by optimizing a variant of variational bound on the data likelihood. DiffWave produces high-fidelity audio in different waveform generation tasks, including neural vocoding conditioned on mel spectrogram, class-conditional generation, and unconditional generation. We demonstrate that DiffWave matches a strong WaveNet vocoder in terms of speech quality (MOS: 4.44 versus 4.43), while synthesizing orders of magnitude faster. In particular, it significantly outperforms autoregressive and GAN-based waveform models in the challenging unconditional generation task in terms of audio quality and sample diversity from various automatic and human evaluations. [1]

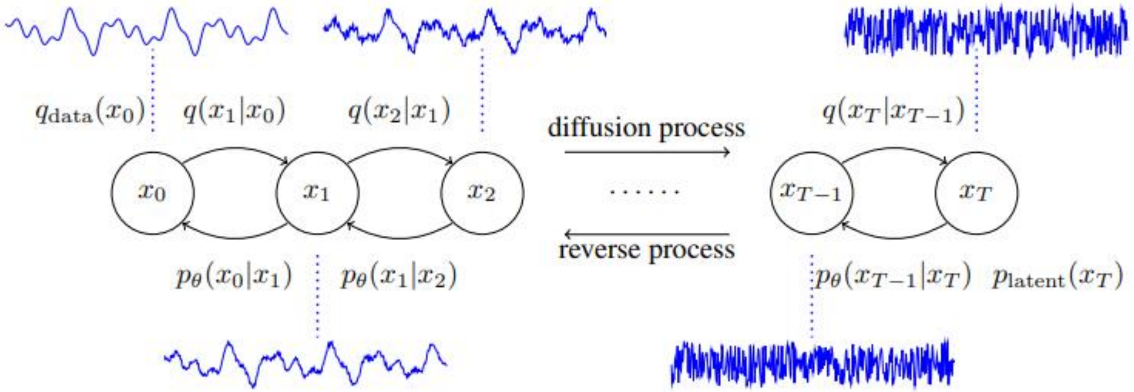
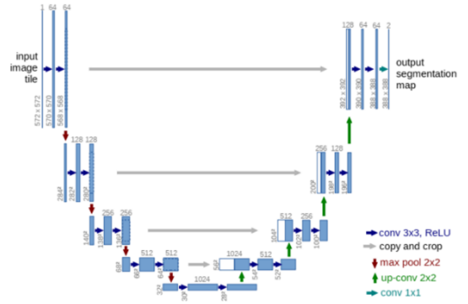


Figure 1: The diffusion and reverse process in diffusion probabilistic models. The reverse process gradually converts the white noise signal into speech waveform through a Markov chain $p_\theta(x_{t-1}|x_t)$.

Kong, Zhifeng, et al. "DiffWave: A Versatile Diffusion Model for Audio Synthesis." International Conference on Learning Representations. 2020.

선행연구

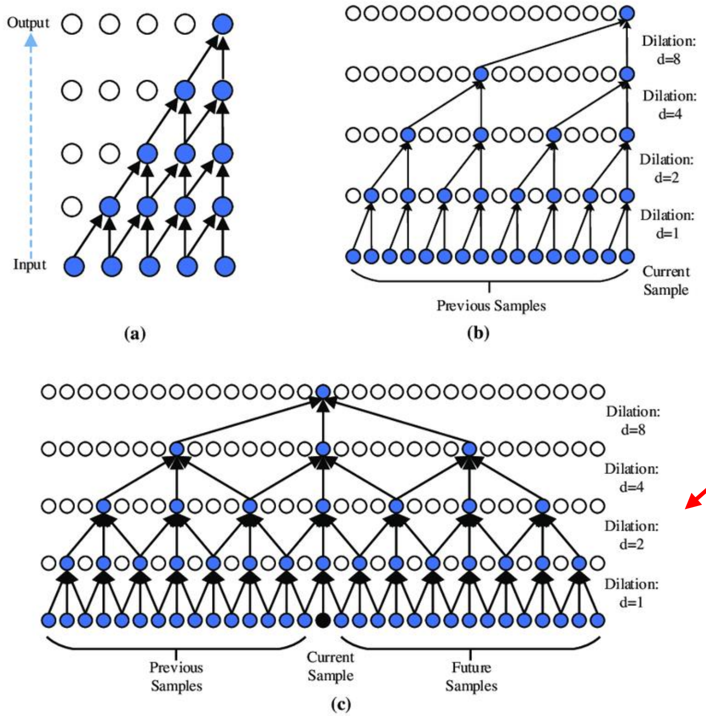
A versatile Diffusion Model for Audio Synthesis

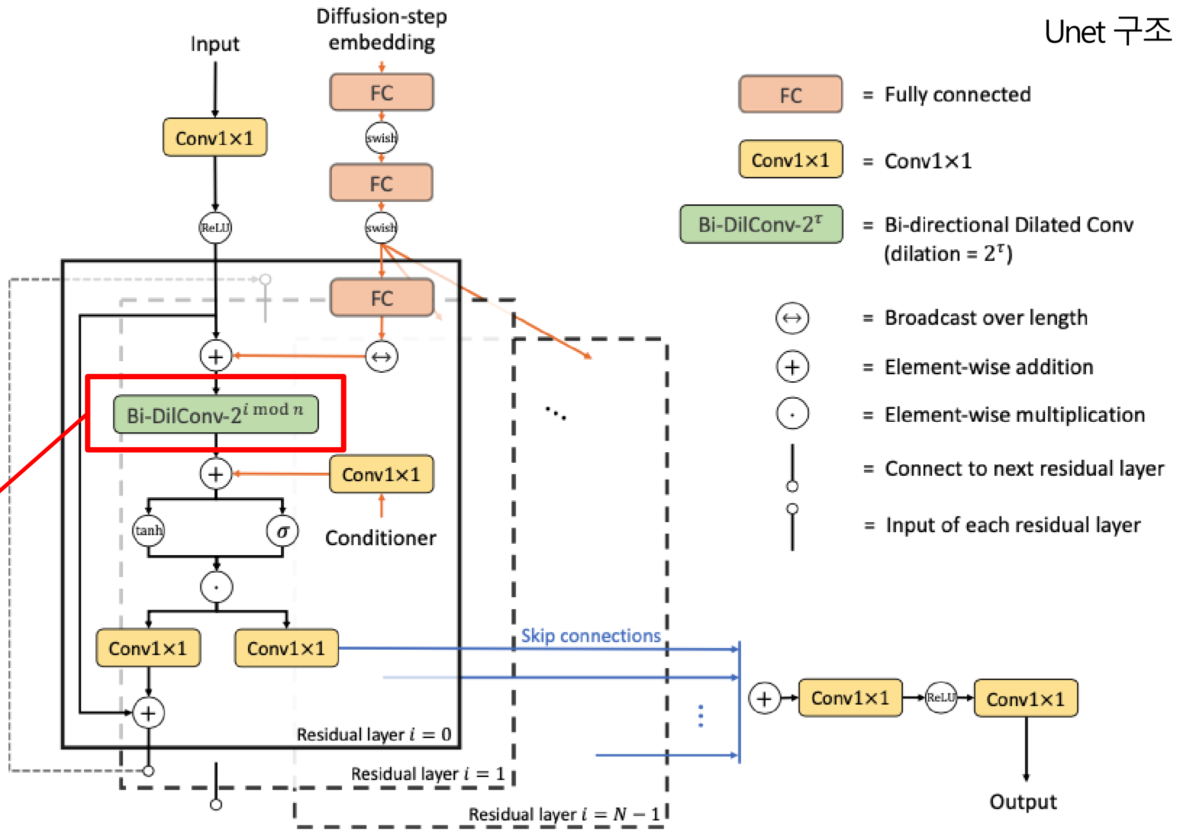


Unet 구조

❖ 어떻게 순차성이 있는 데이터를 다루었을까?

- Inference 단계에서 순차적으로 예측
- Bi-Dilated Convolution Layer – 장기 의존성 반영





Kong, Zhifeng, et al. "DiffWave: A Versatile Diffusion Model for Audio Synthesis." International Conference on Learning Representations. 2020.

Time Series Forecasting

Autoregressive Denoising Diffusion Models for Multivariate Probabilistic Time Series Forecasting

❖ **TimeGrad : Autoregressive Denoising Diffusion Models for Multivariate Probabilistic Time Series Forecasting**

- 2021년에 제안된 Autoregressive diffusion model을 이용한 확률적 다변량 시계열 예측 모델 (ICML, 24년 03월 기준 181회 인용)

Autoregressive Denoising Diffusion Models for Multivariate Probabilistic Time Series Forecasting

---

Kashif Rasul [1] **Calvin Seward** [1] **Ingmar Schuster** [1] **Roland Vollgraf** [1]

Rasul, Kashif, et al. "Autoregressive denoising diffusion models for multivariate probabilistic time series forecasting." International Conference on Machine Learning. PMLR, 2021.

Method

Autoregressive Denoising Diffusion Models for Multivariate Probabilistic Time Series Forecasting

❖ TimeGrad

- 다변량 시계열 예측을 위한 Autoregressive-Diffusion 모델

Data

Time Step	x_1	x_2	…	x_{D-1}	x_D
1	2710.349	1707.4		334	286.898
2	2778.801	1733.7		331	310.859
…					
$t_0 - 1$	2786.545	1765.21		334	296.521
t_0	2753.958	1763.21		321	264.284
…					
$T - 1$	2785.204	1753.7		326	289.226
T	2847.699	1770.5		344	299.356

Method

Autoregressive Denoising Diffusion Models for Multivariate Probabilistic Time Series Forecasting

❖ TimeGrad

- 다변량 시계열 예측을 위한 Autoregressive-Diffusion 모델

Data

Time Step	x_1	x_2	…	x_{D-1}	x_D
1	2710.349	1707.4		334	286.898
2	2778.801	1733.7		331	310.859
…					
t_0-1	2786.545	1765.21		334	296.521
t_0	2753.958	1763.21		321	264.284
…					
$T-1$	2785.204	1753.7		326	289.226
T	2847.699	1770.5		344	299.356

Training data : $x^0_{1:t_0-1}$

Predicting data : $x^0_{t_0:T}$

Conditional Distribution

$$q_x(x^0_{t_0:T} \mid x^0_{1:t_0-1}, c_{1:T}) = \prod_{t=t_0}^{T} q_x(x^0_t \mid x^0_{1:t-1}, c_{1:T})$$

Diffusion step 정보

$$x^{\boxed{0}}_{\boxed{t}}$$

Time step 정보

$$c_{1:T}$$

covariate

Method

Autoregressive Denoising Diffusion Models for Multivariate Probabilistic Time Series Forecasting

❖ TimeGrad

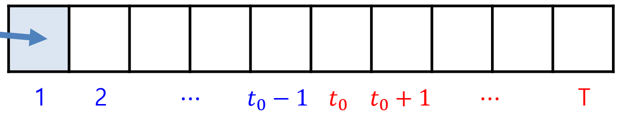
- 다변량 시계열 예측을 위한 Autoregressive-Diffusion 모델

Data

Time Step	x_1	x_2	…	x_{D-1}	x_D
1	2710.349	1707.4		334	286.898
2	2778.801	1733.7		331	310.859
…					
t_0-1	2786.545	1765.21		334	296.521
t_0	2753.958	1763.21		321	264.284
…					
$T-1$	2785.204	1753.7		326	289.226
T	2847.699	1770.5		344	299.356

Conditional Distribution

$$q_x\left(x^0_{t_0:T} \middle| x^0_{1:t_0-1}, c_{1:T}\right) = \prod_{t=t_0}^{T} q_x(x^0_t | x^0_{1:t-1}, c_{1:T})$$



Rasul, K., Seward, C., Schuster, I., & Vollgraf, R. (2021, July). Autoregressive denoising diffusion models for multivariate probabilistic time series forecasting. In International Conference on Machine Learning (pp. 8857-8868). PMLR.

Method

Autoregressive Denoising Diffusion Models for Multivariate Probabilistic Time Series Forecasting

❖ TimeGrad

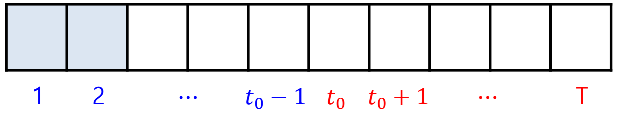
- 다변량 시계열 예측을 위한 Autoregressive-Diffusion 모델

Data

Time Step	x_1	x_2	…	x_{D-1}	x_D
1	2710.349	1707.4		334	286.898
2	2778.801	1733.7		331	310.859
…					
t_0-1	2786.545	1765.21		334	296.521
t_0	2753.958	1763.21		321	264.284
…					
$T-1$	2785.204	1753.7		326	289.226
T	2847.699	1770.5		344	299.356

Conditional Distribution

$$q_x\left(x^0_{t_0:T} \middle| x^0_{1:t_0-1}, c_{1:T}\right) = \prod_{t=t_0}^{T} q_x\left(x^0_t \middle| x^0_{1:t-1}, c_{1:T}\right)$$



Rasul, K., Seward, C., Schuster, I., & Vollgraf, R. (2021, July). Autoregressive denoising diffusion models for multivariate probabilistic time series forecasting. In International Conference on Machine Learning (pp. 8857-8868). PMLR.

Method

Autoregressive Denoising Diffusion Models for Multivariate Probabilistic Time Series Forecasting

❖ TimeGrad

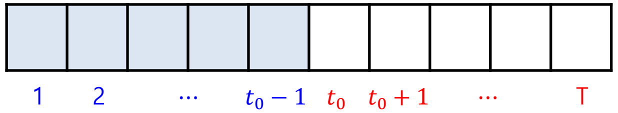
- 다변량 시계열 예측을 위한 Autoregressive-Diffusion 모델

Data

Time Step	x_1	x_2	…	x_{D-1}	x_D
1	2710.349	1707.4		334	286.898
2	2778.801	1733.7		331	310.859
…					
$t_0 - 1$	2786.545	1765.21		334	296.521
t_0	2753.958	1763.21		321	264.284
…					
$T - 1$	2785.204	1753.7		326	289.226
T	2847.699	1770.5		344	299.356

Conditional Distribution

$$q_x\left(x^0_{t_0:T} \middle| x^0_{1:t_0-1}, c_{1:T}\right) = \prod_{t=t_0}^{T} q_x\left(x^0_t \middle| x^0_{1:t-1}, c_{1:T}\right)$$



Rasul, K., Seward, C., Schuster, I., & Vollgraf, R. (2021, July). Autoregressive denoising diffusion models for multivariate probabilistic time series forecasting. In International Conference on Machine Learning (pp. 8857-8868). PMLR.

Method

Autoregressive Denoising Diffusion Models for Multivariate Probabilistic Time Series Forecasting

❖ TimeGrad

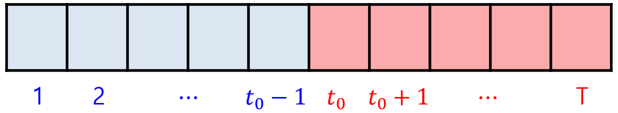
- 다변량 시계열 예측을 위한 Autoregressive-Diffusion 모델

Data

Time Step	x_1	x_2	…	x_{D-1}	x_D
1	2710.349	1707.4		334	286.898
2	2778.801	1733.7		331	310.859
…					
$t_0 - 1$	2786.545	1765.21		334	296.521
t_0	2753.958	1763.21		321	264.284
…					
$T-1$	2785.204	1753.7		326	289.226
T	2847.699	1770.5		344	299.356

Conditional Distribution

$$q_x\left(x^0_{t_0:T} \middle| x^0_{1:t_0-1}, c_{1:T}\right) = \prod_{t=t_0}^{T} q_x\left(x^0_t \middle| x^0_{1:t-1}, c_{1:T}\right)$$



$$q_x\left(x^0_{t_0:T} \middle| x^0_{1:t_0-1}\right)$$

Rasul, K., Seward, C., Schuster, I., & Vollgraf, R. (2021, July). Autoregressive denoising diffusion models for multivariate probabilistic time series forecasting. In International Conference on Machine Learning (pp. 8857-8868). PMLR.

Method

Autoregressive Denoising Diffusion Models for Multivariate Probabilistic Time Series Forecasting

❖ TimeGrad

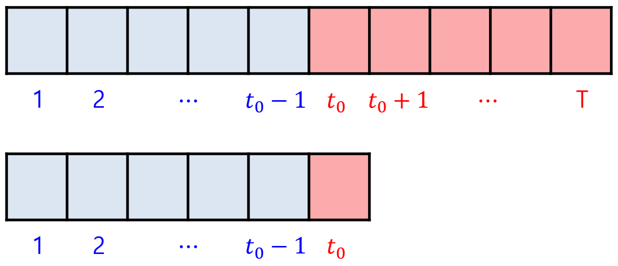
- 다변량 시계열 예측을 위한 Autoregressive-Diffusion 모델

Data

Time Step	$\boldsymbol{x_1}$	$\boldsymbol{x_2}$	…	$\boldsymbol{x_{D-1}}$	$\boldsymbol{x_D}$
1	2710.349	1707.4		334	286.898
2	2778.801	1733.7		331	310.859
…					
t_0-1	2786.545	1765.21		334	296.521
t_0	2753.958	1763.21		321	264.284
…					
$T-1$	2785.204	1753.7		326	289.226
T	2847.699	1770.5		344	299.356

Conditional Distribution

$$q_x\left(x^0_{t_0:T} \middle| x^0_{1:t_0-1}, c_{1:T}\right) = \prod_{t=t_0}^{T} q_x\left(x^0_t \middle| x^0_{1:t-1}, c_{1:T}\right)$$



$$q_x\left(x^0_{t_0:T} \middle| x^0_{1:t_0-1}\right)$$

$$\parallel$$

$$q_x\left(x^0_{t_0} \middle| x^0_{1:t_0-1}\right)$$

Rasul, K., Seward, C., Schuster, I., & Vollgraf, R. (2021, July). Autoregressive denoising diffusion models for multivariate probabilistic time series forecasting. In International Conference on Machine Learning (pp. 8857-8868). PMLR.

Method

Autoregressive Denoising Diffusion Models for Multivariate Probabilistic Time Series Forecasting

❖ TimeGrad

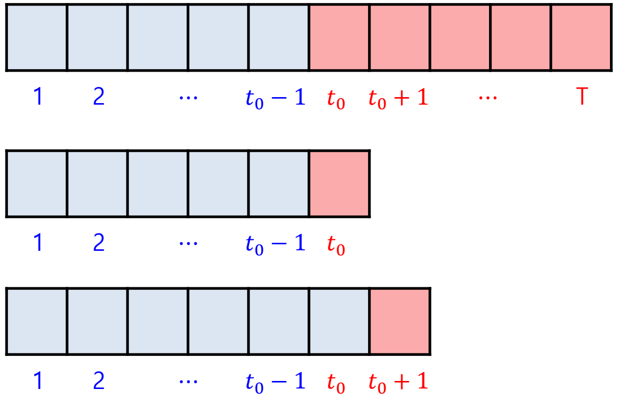
- 다변량 시계열 예측을 위한 Autoregressive-Diffusion 모델

Data

Time Step	x_1	x_2	…	x_{D-1}	x_D
1	2710.349	1707.4		334	286.898
2	2778.801	1733.7		331	310.859
…					
$t_0 - 1$	2786.545	1765.21		334	296.521
t_0	2753.958	1763.21		321	264.284
…					
$T - 1$	2785.204	1753.7		326	289.226
T	2847.699	1770.5		344	299.356

Conditional Distribution

$$q_x\left(x^0_{t_0:T} \middle| x^0_{1:t_0-1}, c_{1:T}\right) = \prod_{t=t_0}^{T} q_x\left(x^0_t \middle| x^0_{1:t-1}, c_{1:T}\right)$$



$$q_x\left(x^0_{t_0:T} \middle| x^0_{1:t_0-1}\right)$$
$$\parallel$$
$$q_x\left(x^0_{t_0} \middle| x^0_{1:t_0-1}\right)$$
$$\times$$
$$q_x\left(x^0_{t_0+1} \middle| x^0_{1:t_0}\right)$$

Rasul, K., Seward, C., Schuster, I., & Vollgraf, R. (2021, July). Autoregressive denoising diffusion models for multivariate probabilistic time series forecasting. In International Conference on Machine Learning (pp. 8857-8868). PMLR.

Method

Autoregressive Denoising Diffusion Models for Multivariate Probabilistic Time Series Forecasting

❖ TimeGrad

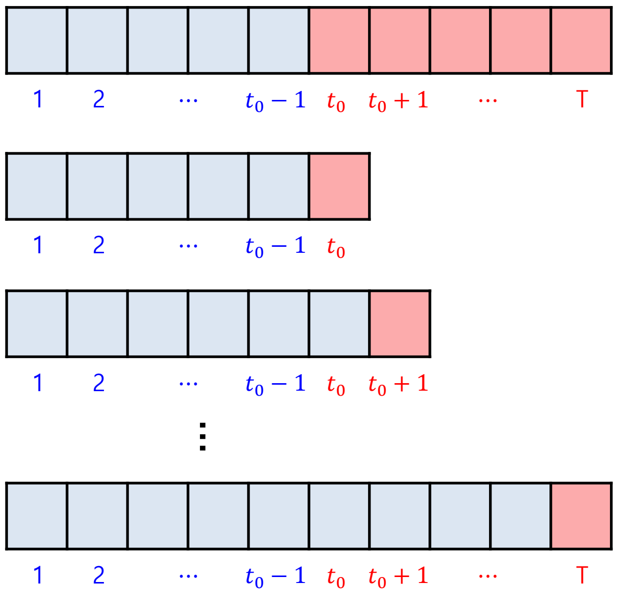
- 다변량 시계열 예측을 위한 Autoregressive-Diffusion 모델

Data

Time Step	x_1	x_2	…	x_{D-1}	x_D
1	2710.349	1707.4		334	286.898
2	2778.801	1733.7		331	310.859
…					
$t_0 - 1$	2786.545	1765.21		334	296.521
t_0	2753.958	1763.21		321	264.284
…					
$T - 1$	2785.204	1753.7		326	289.226
T	2847.699	1770.5		344	299.356

Conditional Distribution

$$q_x\left(x^0_{t_0:T} \middle| x^0_{1:t_0-1}, c_{1:T}\right) = \prod_{t=t_0}^{T} q_x\left(x^0_t \middle| x^0_{1:t-1}, c_{1:T}\right)$$

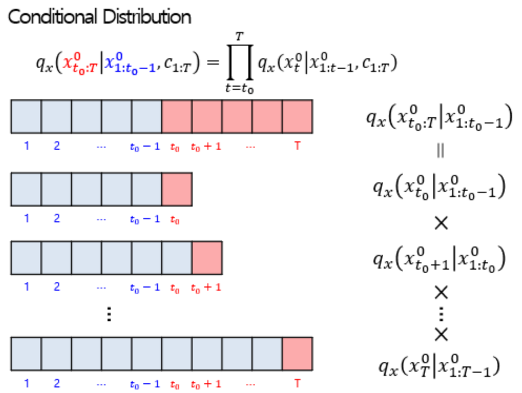


$$q_x\left(x^0_{t_0:T} \middle| x^0_{1:t_0-1}\right)$$
$$\|$$
$$q_x\left(x^0_{t_0} \middle| x^0_{1:t_0-1}\right)$$
$$\times$$
$$q_x\left(x^0_{t_0+1} \middle| x^0_{1:t_0}\right)$$
$$\times$$
$$\vdots$$
$$\times$$
$$q_x\left(x^0_{T} \middle| x^0_{1:T-1}\right)$$

Rasul, K., Seward, C., Schuster, I., & Vollgraf, R. (2021, July). Autoregressive denoising diffusion models for multivariate probabilistic time series forecasting. In International Conference on Machine Learning (pp. 8857-8868). PMLR.

Method

Autoregressive Denoising Diffusion Models for Multivariate Probabilistic Time Series Forecasting



❖ Method

- 매 timestep마다 diffusion process 진행하여 모델을 학습
- 이전 시점의 정보를 반영하기 위해 RNN 사용

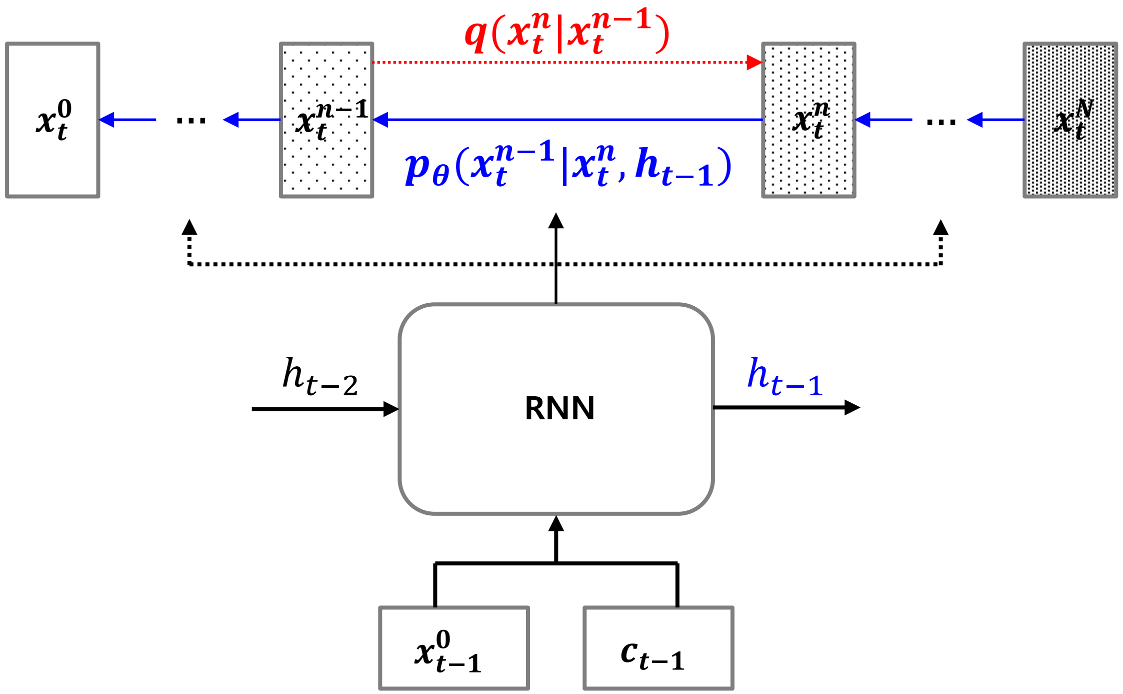
Conditional Distribution

$$q_x\left(x^0_{t_0:T}\middle|x^0_{1:t_0-1}, c_{1:T}\right) = \prod_{t=t_0}^{T} q_x(x^0_t|x^0_{1:t-1}, c_{1:T})$$

$$\rightarrow \prod_{t=t_0}^{T} p_\theta(x^0_t|h_{t-1})$$

$$h_t = \mathrm{RNN}_\theta(concat(x^0_t, c_t), h_{t-1})$$

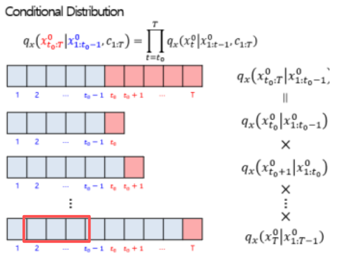
RNN conditioned diffusion probabilistic model

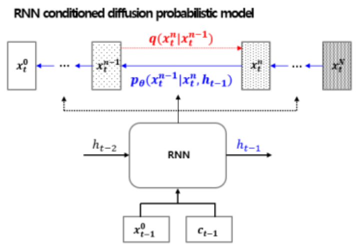


Rasul, K., Seward, C., Schuster, I., & Vollgraf, R. (2021, July). Autoregressive denoising diffusion models for multivariate probabilistic time series forecasting. In International Conference on Machine Learning (pp. 8857-8868). PMLR.

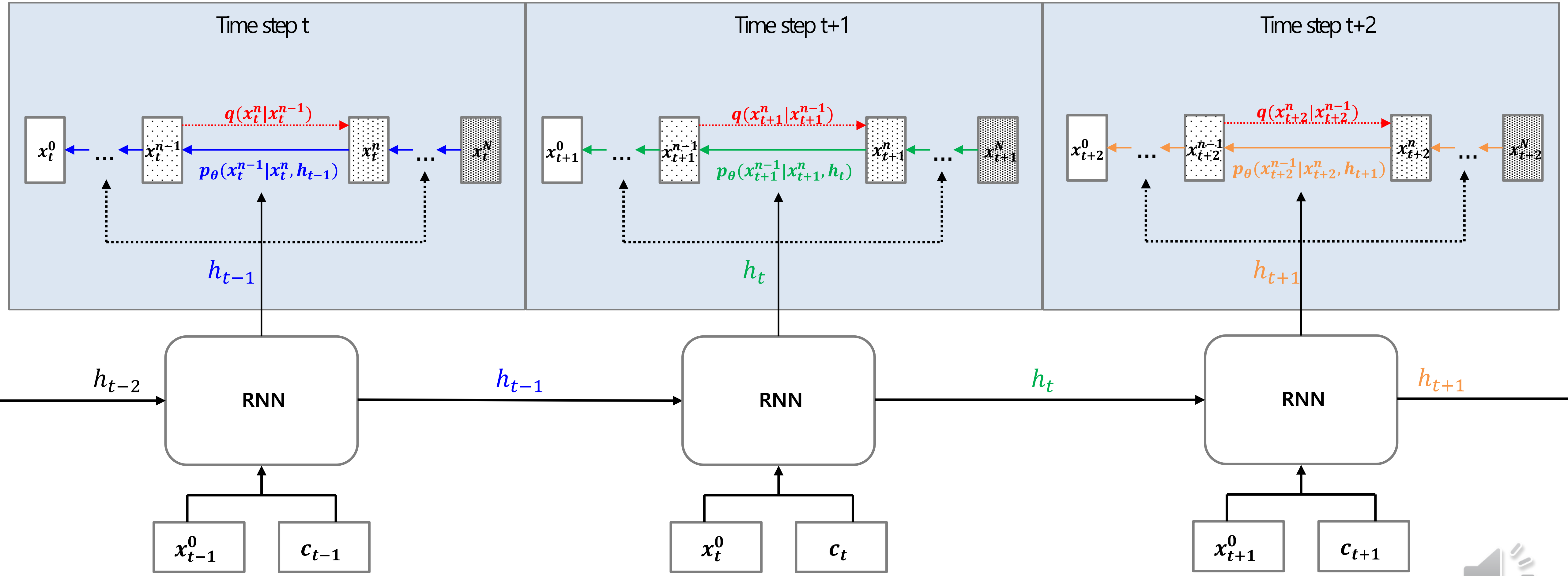
Method

Autoregressive Denoising Diffusion Models for Multivariate Probabilistic Time Series Forecasting





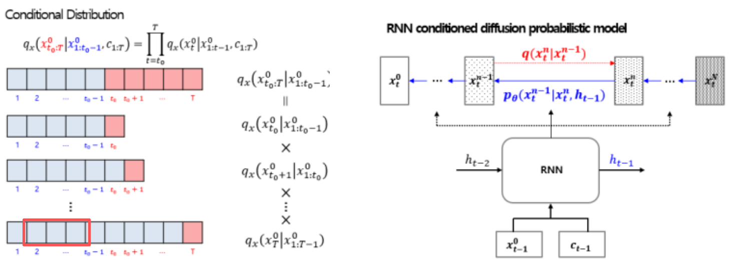
❖ Training Process total



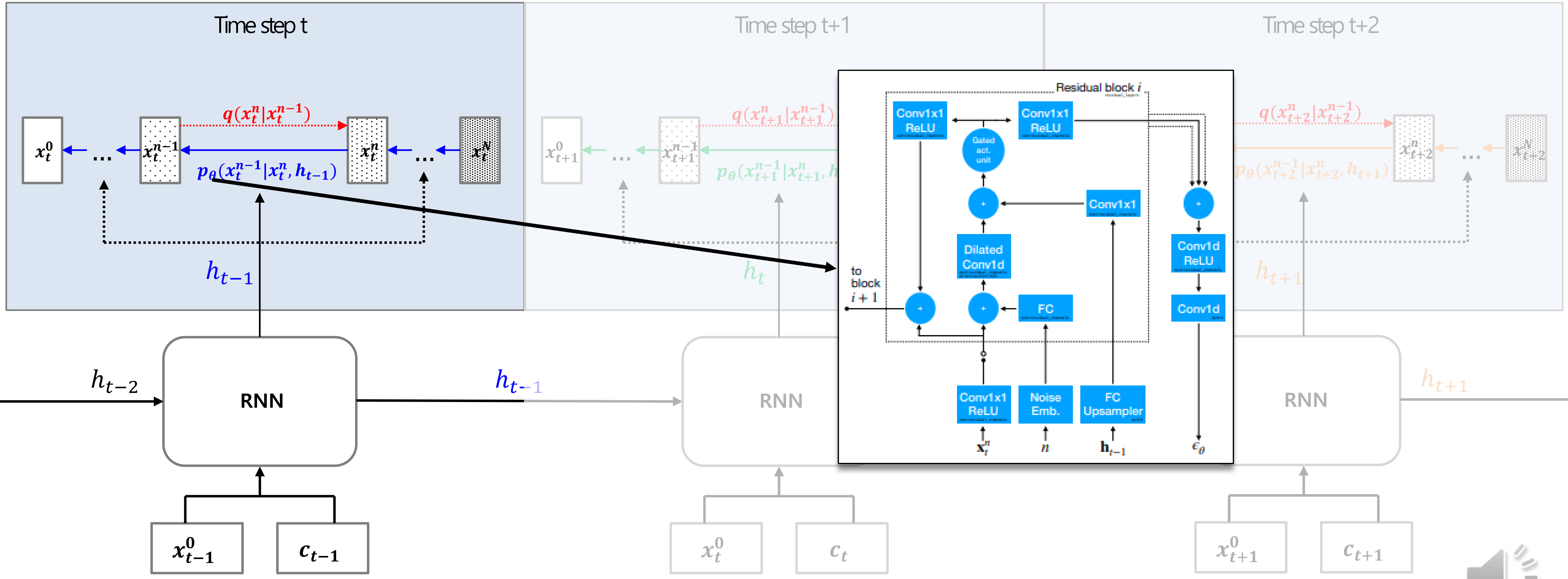
Rasul, K., Seward, C., Schuster, I., & Vollgraf, R. (2021, July). Autoregressive denoising diffusion models for multivariate probabilistic time series forecasting. In International Conference on Machine Learning (pp. 8857-8868). PMLR.

Method

Autoregressive Denoising Diffusion Models for Multivariate Probabilistic Time Series Forecasting



❖ **Training Process total**



Rasul, K., Seward, C., Schuster, I., & Vollgraf, R. (2021, July). Autoregressive denoising diffusion models for multivariate probabilistic time series forecasting. In International Conference on Machine Learning (pp. 8857-8868). PMLR.

Experiments

Autoregressive Denoising Diffusion Models for Multivariate Probabilistic Time Series Forecasting

❖ Experiments

- ✓ Adam (learning rate 0.001)
- ✓ $\beta_1(0.0001) \rightarrow \beta_N(0.1)$
- ✓ Diffusion step (N=100)

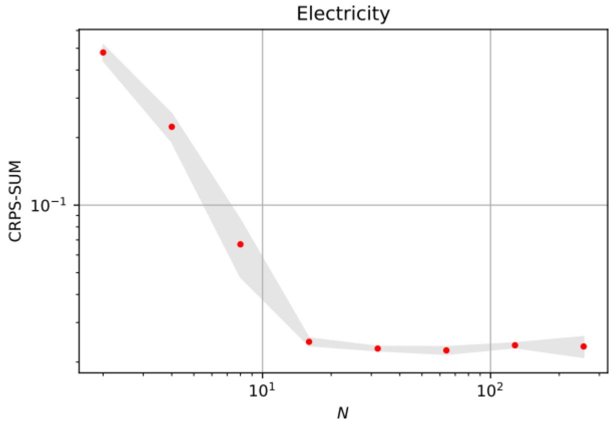


Table 2. Test set CRPS_{sum} comparison (lower is better) of models on six real world data sets. Mean and standard error metrics for `TimeGrad` obtained by re-training and evaluating 10 times.

Method	Exchange	Solar	Electricity	Traffic	Taxi	Wikipedia
VES	**0.005**±0.000	0.9±0.003	0.88±0.0035	0.35±0.0023	-	-
VAR	**0.005**±0.000	0.83±0.006	0.039±0.0005	0.29±0.005	-	-
VAR-Lasso	0.012±0.0002	0.51±0.006	0.025±0.0002	0.15±0.002	-	3.1±0.004
GARCH	0.023±0.000	0.88±0.002	0.19±0.001	0.37±0.0016	-	-
KVAE	0.014±0.002	0.34±0.025	0.051±0.019	0.1±0.005	-	0.095±0.012
Vec-LSTM ind-scaling	0.008±0.001	0.391±0.017	0.025±0.001	0.087±0.041	0.506±0.005	0.133±0.002
Vec-LSTM lowrank-Copula	0.007±0.000	0.319±0.011	0.064±0.008	0.103±0.006	0.326±0.007	0.241±0.033
GP scaling	0.009±0.000	0.368±0.012	0.022±0.000	0.079±0.000	0.183±0.395	1.483±1.034
GP Copula	0.007±0.000	0.337±0.024	0.0245±0.002	0.078±0.002	0.208±0.183	0.086±0.004
Transformer MAF	**0.005**±0.003	0.301±0.014	0.0207±0.000	0.056±0.001	0.179±0.002	0.063±0.003
TimeGrad	0.006±0.001	**0.287**±0.02	**0.0206**±0.001	**0.044**±0.006	**0.114**±0.02	**0.0485**±0.002

Rasul, K., Seward, C., Schuster, I., & Vollgraf, R. (2021, July). Autoregressive denoising diffusion models for multivariate probabilistic time series forecasting. In International Conference on Machine Learning (pp. 8857-8868). PMLR.

Summary

Autoregressive Denoising Diffusion Models for Multivariate Probabilistic Time Series Forecasting

❖ **TimeGrad**

- Autoregressive Model과 Diffusion Model을 활용한 다변량 확률적 시계열 예측
- 뛰어난 성능을 보임

❖ **Limitation**

- 샘플링 과정 많은 시간 소요
- 데이터 결측치 처리를 할 수 없음

Rasul, K., Seward, C., Schuster, I., & Vollgraf, R. (2021, July). Autoregressive denoising diffusion models for multivariate probabilistic time series forecasting. In International Conference on Machine Learning (pp. 8857-8868). PMLR.

Time Series Imputation

Time Series Missing Data

❖ 시계열 예측 Challenges

- 적합한 확률 분포 모델 필요
- 숨겨진 여러 형태의 패턴을 찾아 학습
- 다변량 시계열 데이터 다루기
- 노이즈 또는 결측치 처리

❖ Time Series Missing Data

- 현실에서 접하는 대부분 데이터는 항상 결측치 발생
  - ✓ 시간의 흐름에 따라 그 시기에 데이터를 직접 수집하는 시계열에서 발생 빈도가 높음
- 특정 시점의 평균과 분산의 왜곡을 가져오게 되어 **Critical**

MCAR

완전 무작위 결측
(Missing Completely At Random)

- 결측 발생이 다른 변수와 상관이 없는 경우

예) 사람이 혈액 샘플을 담고 있는 관이 우연히 떨어지고 파손된 경우

$$f(M|Y,\emptyset) = f(M|\emptyset) for\ all\ Y,\emptyset$$

MAR

무작위 결측
(Missing At Random)

- 결측 발생이 특정 변수와 연관이 있으나 얻고자 하는 결과와 상관이 없는 경우

예) 설문에서 고객의 소득 수준이 빠졌지만, 고객의 직업, 경험 및 자격과 같은 다른 변수에서 추정 가능

$$f(M|Y,\emptyset) = f(M|Y_{obs},\emptyset) for\ all\ Y_{mis},\emptyset$$

MNAR

비무작위 결측
(Missing Not At Random)

- 결측 발생이 다른 변수와 상관이 있는 경우

MCAR 혹은 MAR이 아닌 경우

$$f(M,Y|\theta,\emptyset) = f(Y|\theta)f(M|Y,\emptyset)$$

Osman, M. S., Abu-Mahfouz, A. M., & Page, P. R. (2018). A survey on data imputation techniques: Water distribution system as a use case. IEEE Access, 6, 63279-63291.

Time Series Imputation

Conditional Score-Based Diffusion Models for Probabilistic Time Series Imputation

❖ Conditional Score-Based Diffusion Models for Probabilistic Time Series Imputation

- 2021년에 제안된 Conditional Score-based Diffusion Model을 이용한 Time Series Imputation model (NuerIPS, 24년 03월 기준 231회 인용)

CSDI: Conditional Score-based Diffusion Models for Probabilistic Time Series Imputation

Yusuke Tashiro[1,2,3*], Jiaming Song[1], Yang Song[1], Stefano Ermon[1]
[1]Department of Computer Science, Stanford University, Stanford, CA, USA
[2]Mitsubishi UFJ Trust Investment Technology Institute, Tokyo, Japan
[3]Japan Digital Design, Tokyo, Japan
{ytashiro,tsong,songyang,ermon}@cs.stanford.edu

Tashiro, Yusuke, et al. "Csdi: Conditional score-based diffusion models for probabilistic time series imputation." Advances in Neural Information Processing Systems 34 (2021): 24804-24816.

Goal

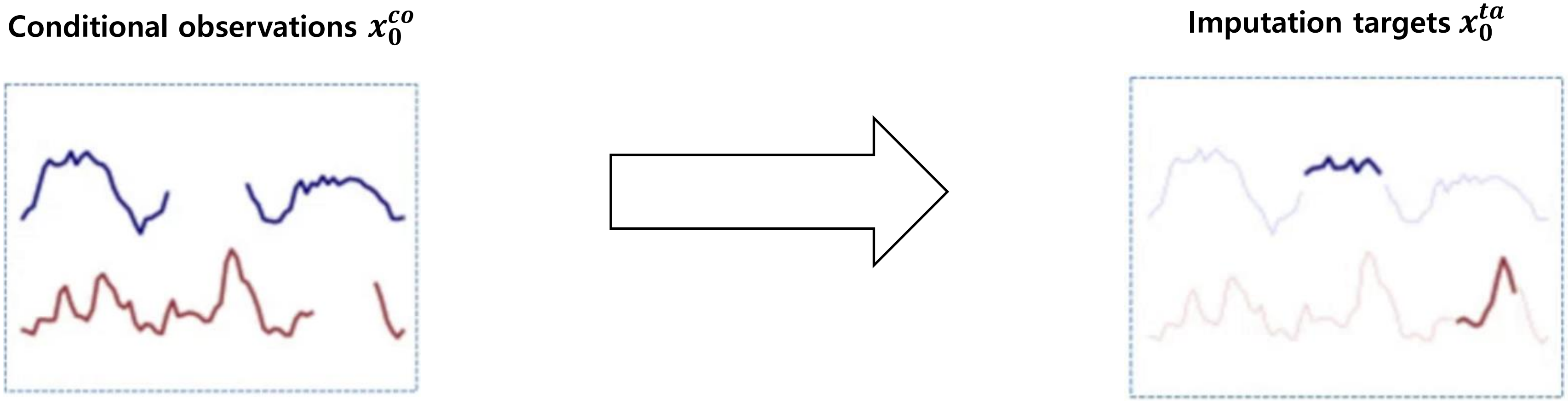
Conditional Score-Based Diffusion Models for Probabilistic Time Series Imputation

❖ Goal of Probabilistic Imputation

- Model distribution으로 True conditional distribution 구하기
- $p_\theta(x_0^{ta}|x_0^{co})$: Model distribution
- $q(x_0^{ta}|x_0^{co})$: True conditional distribution

❖ Imputation Task

- x_0 : Missing Value 포함
- x_0^{co} : Conditional Observations
- x_0^{ta} : Imputation Targets



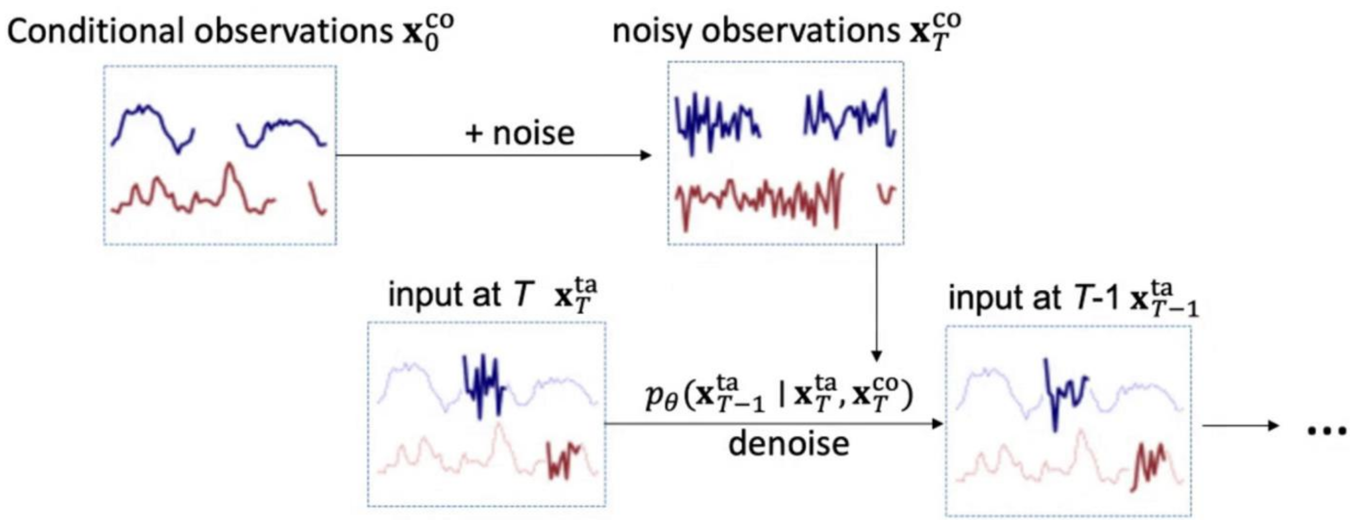
Tashiro, Yusuke, et al. "Csdi: Conditional score-based diffusion models for probabilistic time series imputation." Advances in Neural Information Processing Systems 34 (2021): 24804-24816.

Previous Method

Conditional Score-Based Diffusion Models for Probabilistic Time Series Imputation

❖ Previous Imputation

- Conditional observation & Target imputations에 노이즈 추가



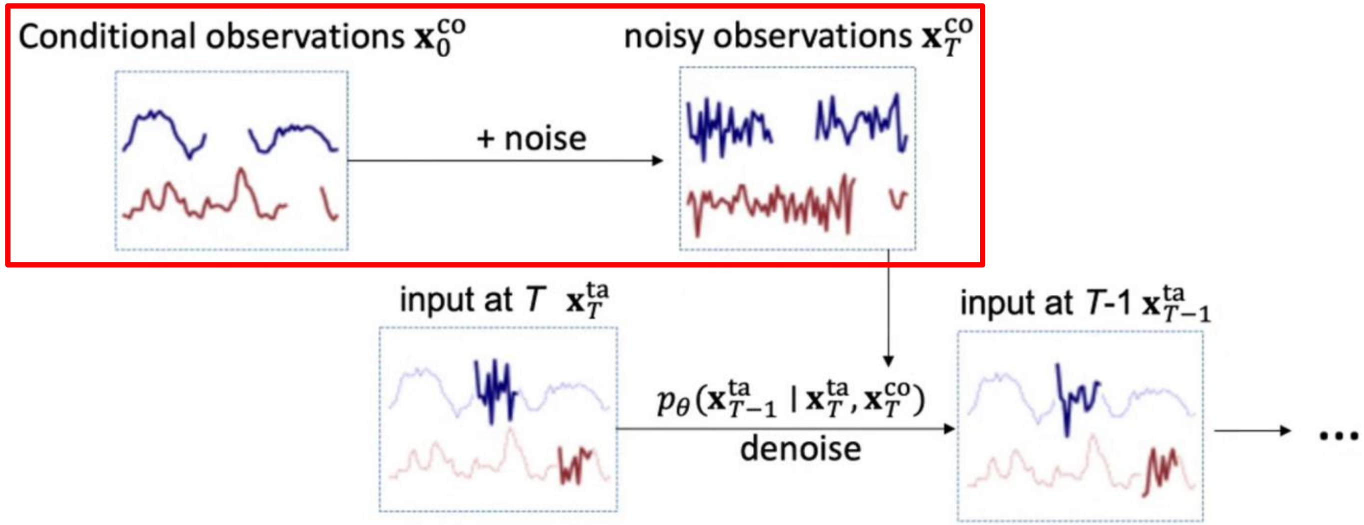
Tashiro, Yusuke, et al. "Csdi: Conditional score-based diffusion models for probabilistic time series imputation." Advances in Neural Information Processing Systems 34 (2021): 24804-24816.

Previous Method

Conditional Score-Based Diffusion Models for Probabilistic Time Series Imputation

❖ Previous Imputation

- Conditional observation & Target imputations에 노이즈 추가



추가된 노이즈에 정보손실이 발생할 우려가 큼

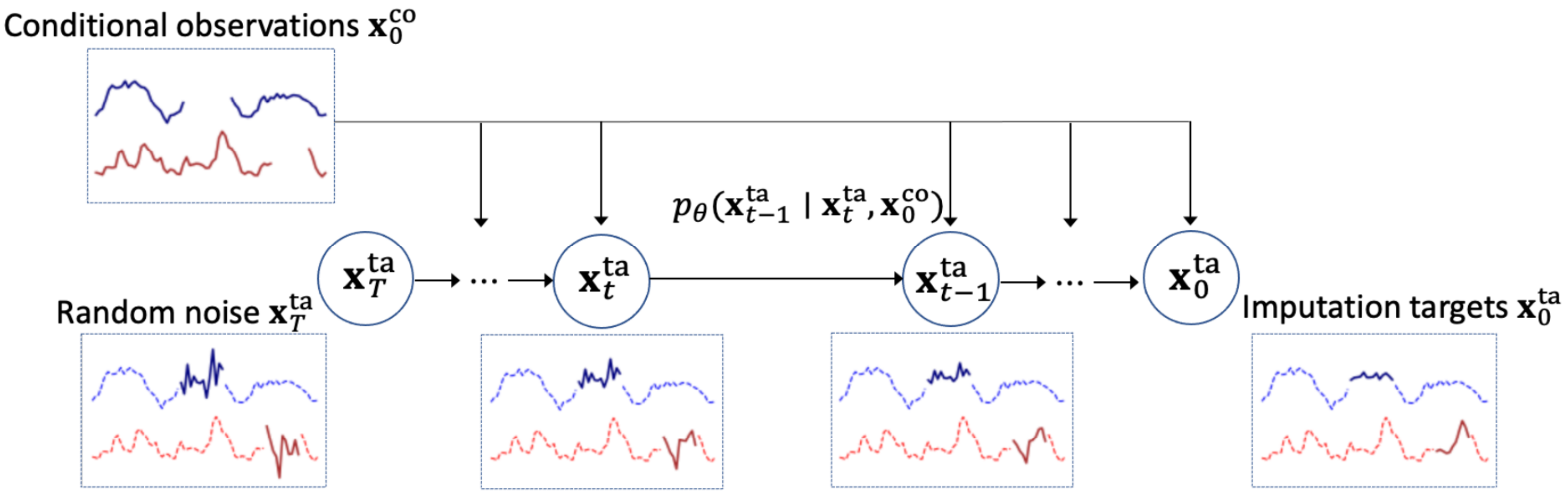
Tashiro, Yusuke, et al. "Csdi: Conditional score-based diffusion models for probabilistic time series imputation." Advances in Neural Information Processing Systems 34 (2021): 24804-24816.

Method

Conditional Score-Based Diffusion Models for Probabilistic Time Series Imputation

❖ Conditional Score-Based Diffusion Models for Probabilistic Time Series Imputation - CSDI

- 관측된 값을 conditional input으로 사용



Tashiro, Yusuke, et al. "Csdi: Conditional score-based diffusion models for probabilistic time series imputation." Advances in Neural Information Processing Systems 34 (2021): 24804-24816.

Method

Conditional Score-Based Diffusion Models for Probabilistic Time Series Imputation

❖ CSDI training method

- Inspired by masked language modeling – self-supervised training method

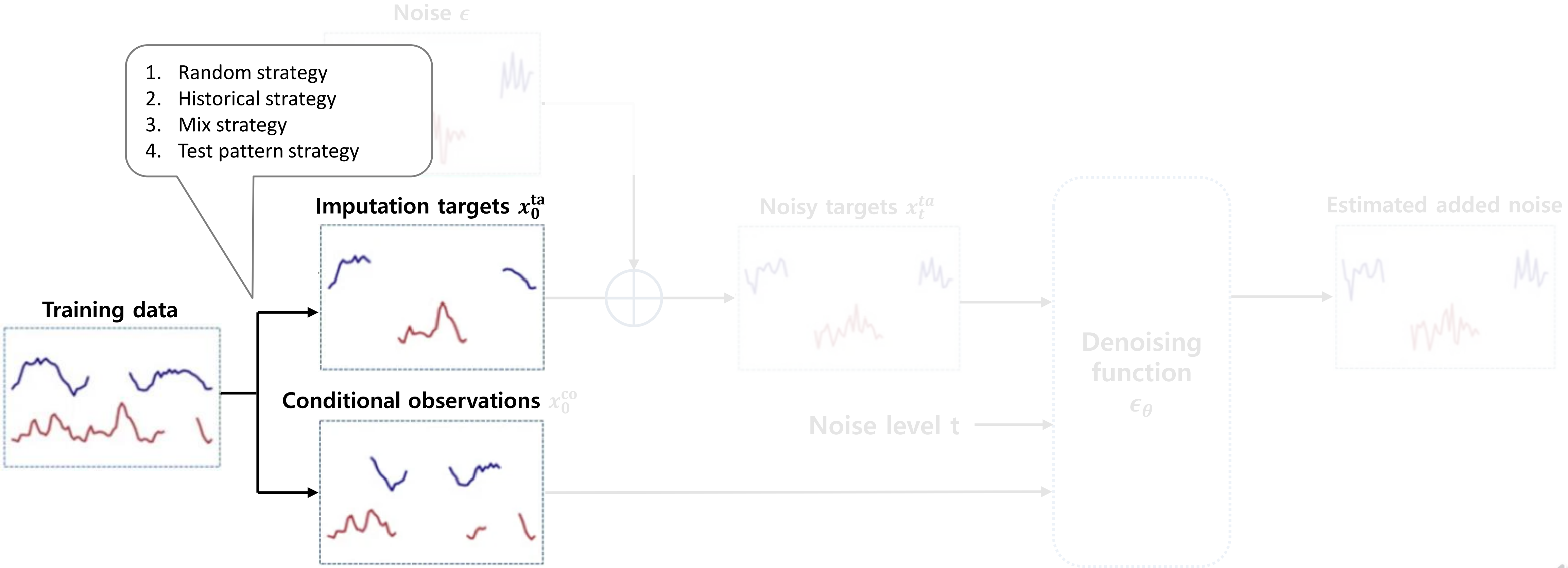
Tashiro, Yusuke, et al. "Csdi: Conditional score-based diffusion models for probabilistic time series imputation." Advances in Neural Information Processing Systems 34 (2021): 24804-24816.
https://papertalk.org/papertalks/37403

Method

Conditional Score-Based Diffusion Models for Probabilistic Time Series Imputation

❖ CSDI training method

- Inspired by masked language modeling – self-supervised training method



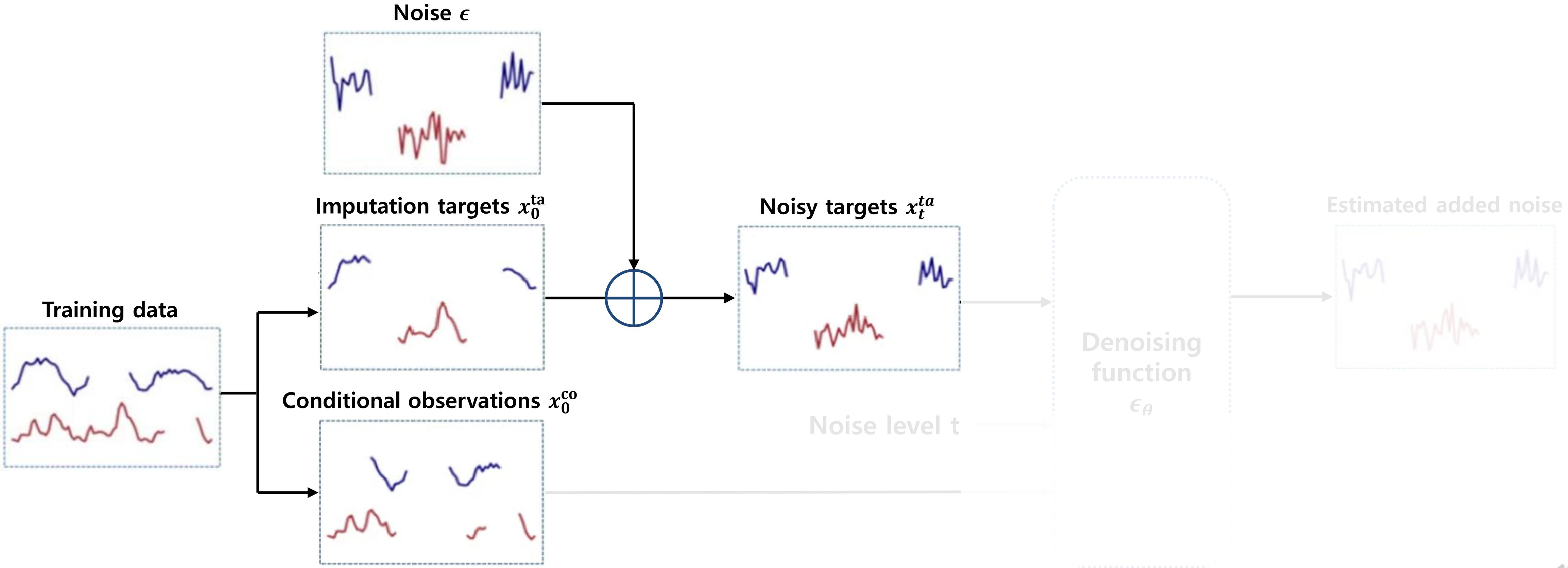
Tashiro, Yusuke, et al. "Csdi: Conditional score-based diffusion models for probabilistic time series imputation." Advances in Neural Information Processing Systems 34 (2021): 24804-24816.
https://papertalk.org/papertalks/37403

Method

Conditional Score-Based Diffusion Models for Probabilistic Time Series Imputation

❖ CSDI training method

- Inspired by masked language modeling – self-supervised training method



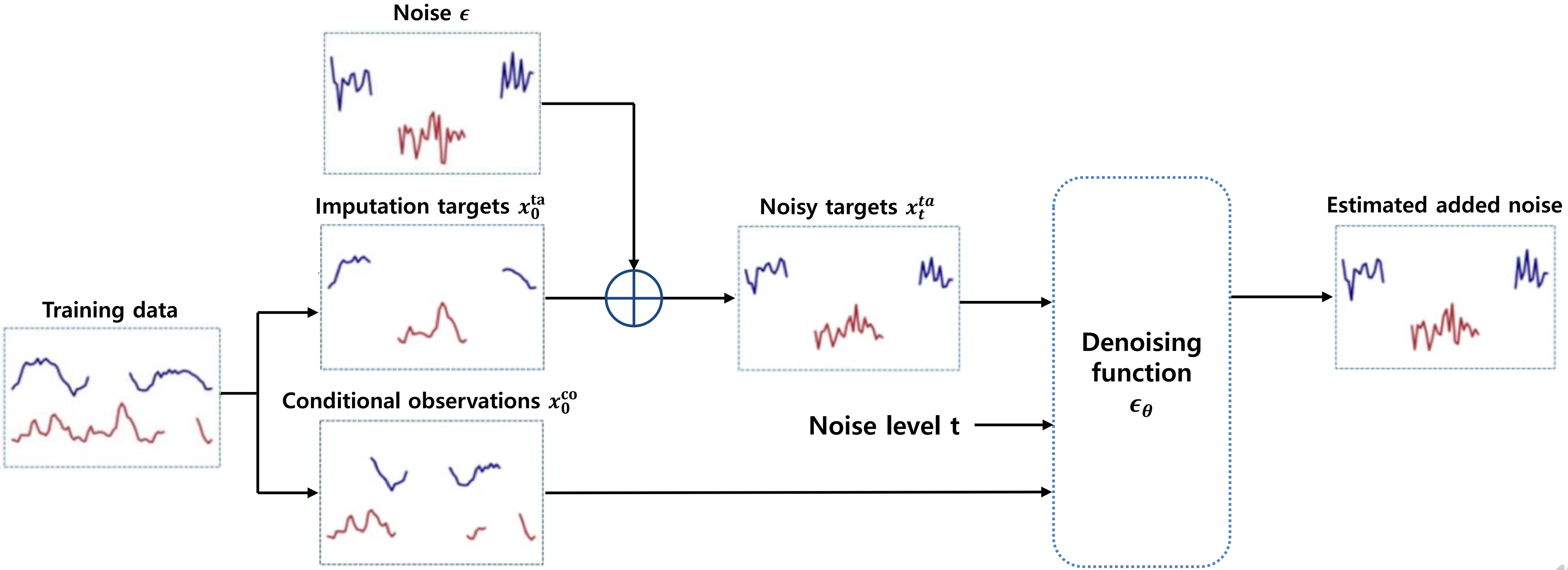
Tashiro, Yusuke, et al. "Csdi: Conditional score-based diffusion models for probabilistic time series imputation." Advances in Neural Information Processing Systems 34 (2021): 24804-24816.
https://papertalk.org/papertalks/37403

Method

Conditional Score-Based Diffusion Models for Probabilistic Time Series Imputation

❖ CSDI training method

- Inspired by masked language modeling – self-supervised training method



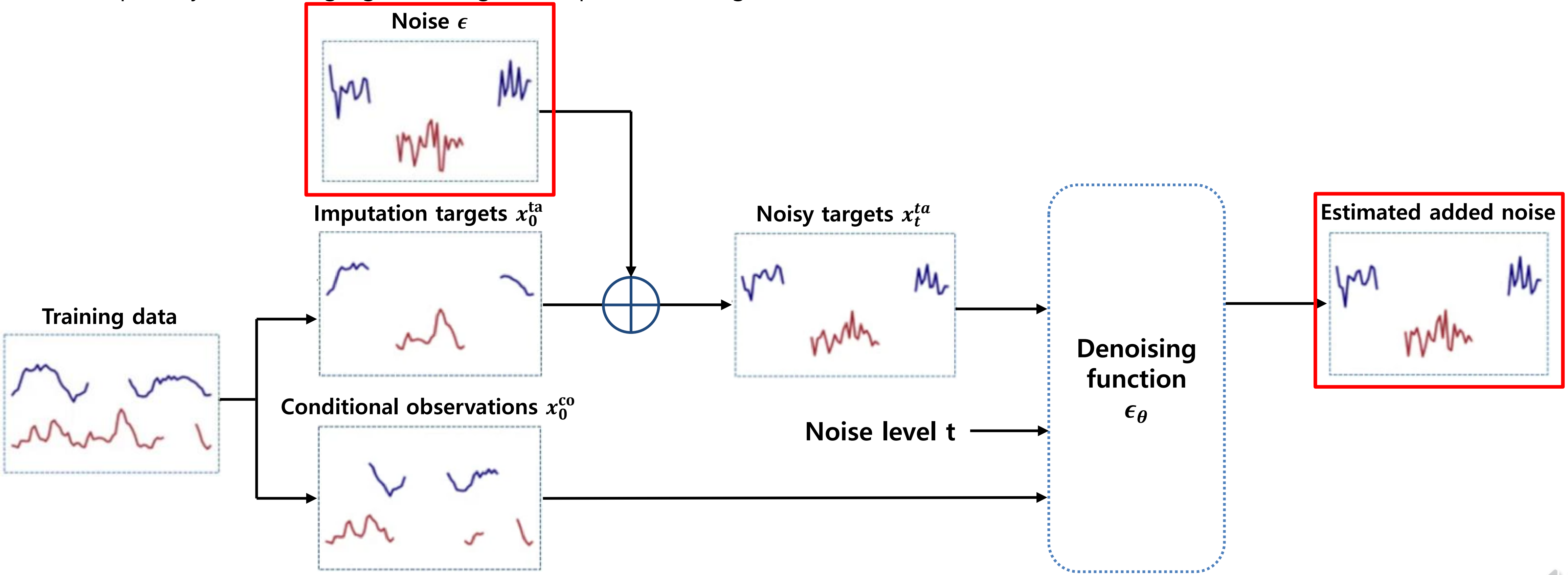
Tashiro, Yusuke, et al. "Csdi: Conditional score-based diffusion models for probabilistic time series imputation." Advances in Neural Information Processing Systems 34 (2021): 24804-24816.
https://papertalk.org/papertalks/37403

Method

Conditional Score-Based Diffusion Models for Probabilistic Time Series Imputation

❖ CSDI training method

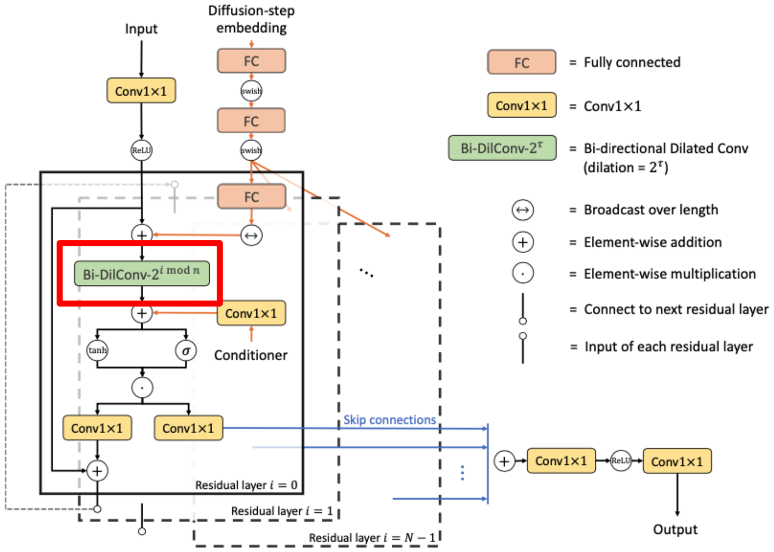
- Inspired by masked language modeling – self-supervised training method



Tashiro, Yusuke, et al. "Csdi: Conditional score-based diffusion models for probabilistic time series imputation." Advances in Neural Information Processing Systems 34 (2021): 24804-24816.
https://papertalk.org/papertalks/37403

Method

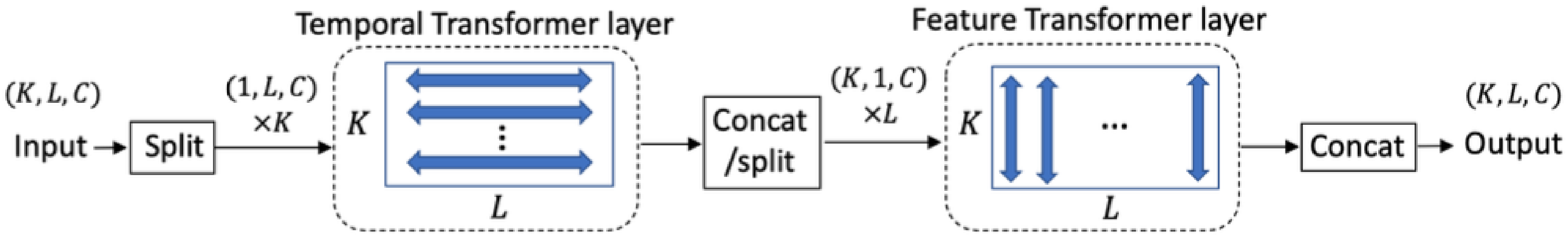
Conditional Score-Based Diffusion Models for Probabilistic Time Series Imputation



❖ 특징

- 2D attention mechanism
  - ✓ 시간 의존성(Time Dependency) 반영 – Temporal Transformer layer
  - ✓ 특성 의존성(Feature Dependency) 반영 – Feature Transformer layer

K : Features L : Length C : Channel



Tashiro, Yusuke, et al. "Csdi: Conditional score-based diffusion models for probabilistic time series imputation." Advances in Neural Information Processing Systems 34 (2021): 24804-24816.

Experiments

Conditional Score-Based Diffusion Models for Probabilistic Time Series Imputation

❖ Dataset

- Health care dataset – PhysioNet Challenge 2012
  - ✓ 4000 clinical time series
  - ✓ 35 변수, 집중 치료실의 48시간
- Air quality dataset
  - ✓ 12개월동안 36개 측정소에서 수집
  - ✓ 시간당 샘플링 된 PM2.5 측정치 사용

Tashiro, Yusuke, et al. "Csdi: Conditional score-based diffusion models for probabilistic time series imputation." Advances in Neural Information Processing Systems 34 (2021): 24804-24816.

Experiments

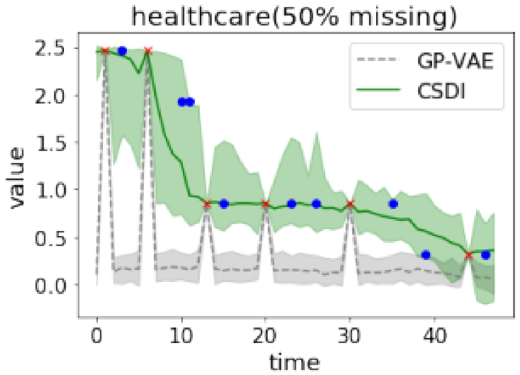
Conditional Score-Based Diffusion Models for Probabilistic Time Series Imputation

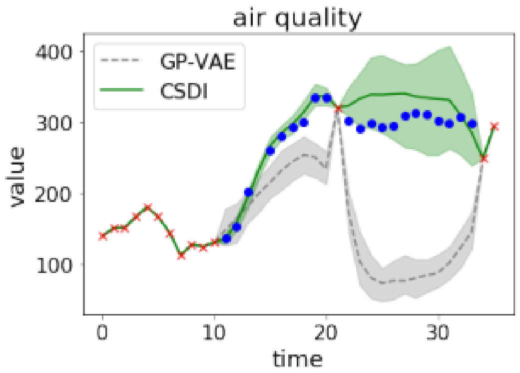
❖ Experiments

- Probabilistic imputation

Table 2: Comparing CRPS for probabilistic imputation baselines and CSDI (lower is better). We report the mean and the standard error of CRPS for five trials.

	healthcare			air quality
	10% missing	50% missing	90% missing	
Multitask GP [31]	0.489(0.005)	0.581(0.003)	0.942(0.010)	0.301(0.003)
GP-VAE [10]	0.574(0.003)	0.774(0.004)	0.998(0.001)	0.397(0.009)
V-RIN [32]	0.808(0.008)	0.831(0.005)	0.922(0.003)	0.526(0.025)
unconditional	0.360(0.007)	0.458(0.008)	0.671(0.007)	0.135(0.001)
CSDI (proposed)	**0.238**(**0.001**)	**0.330**(**0.002**)	**0.522**(**0.002**)	**0.108**(**0.001**)





Tashiro, Yusuke, et al. "Csdi: Conditional score-based diffusion models for probabilistic time series imputation." Advances in Neural Information Processing Systems 34 (2021): 24804-24816.

Experiments

Conditional Score-Based Diffusion Models for Probabilistic Time Series Imputation

❖ Experiments

- Deterministic imputation

Table 3: Comparing MAE for deterministic imputation methods and CSDI. We report the mean and the standard error for five trials. The asterisks mean the results of the method are cited from the original paper.

	healthcare			air quality
	10% missing	50% missing	90% missing	
V-RIN [32]	0.271(0.001)	0.365(0.002)	0.606(0.006)	25.4(0.62)
BRITS [7]	0.284(0.001)	0.368(0.002)	0.517(0.002)	14.11(0.26)
BRITS [7] (*)	0.278	–	–	11.56
GLIMA [21] (*)	0.265	–	–	10.54
RDIS [20]	0.319(0.002)	0.419(0.002)	0.631(0.002)	22.11(0.35)
unconditional	0.326(0.008)	0.417(0.010)	0.625(0.010)	12.13(0.07)
CSDI (proposed)	**0.217**(**0.001**)	**0.301**(**0.002**)	**0.481**(**0.003**)	**9.60**(**0.04**)

Tashiro, Yusuke, et al. "Csdi: Conditional score-based diffusion models for probabilistic time series imputation." Advances in Neural Information Processing Systems 34 (2021): 24804-24816.

Summary

Conditional Score-Based Diffusion Models for Probabilistic Time Series Imputation

❖ CSDI

- Conditional diffusion model을 이용한 Time Series Imputation model
- 다른 Probabilistic & Deterministic Imputation models보다 우수한 성능

❖ Limitation

- 샘플링 속도가 느림
- 장기의존성을 잡기 위한 Computation Cost가 높음

Rasul, K., Seward, C., Schuster, I., & Vollgraf, R. (2021, July). Autoregressive denoising diffusion models for multivariate probabilistic time series forecasting. In International Conference on Machine Learning (pp. 8857-8868). PMLR.

SSSD

Diffusion-based Time Series Imputation and Forecasting with Structured State Space Models

❖ **Diffusion-based Time Series Imputation and Forecasting with Structured State Space Model**

- 2022년 제안된 Diffusion과 Structured State Space Model을 활용한 Time Series Imputation Model (TMLR, 24년 3월 기준 71회 인용)

Diffusion-based Time Series Imputation and Forecasting with Structured State Space Models

Juan Miguel Lopez Alcaraz *juan.lopez.alcaraz@uol.de*
Division AI4Health
Oldenburg University

Nils Strodthoff *nils.strodthoff@uol.de*
Division AI4Health
Oldenburg University

Abstract

The imputation of missing values represents a significant obstacle for many real-world data analysis pipelines. Here, we focus on time series data and put forward SSSD, an imputation model that relies on two emerging technologies, (conditional) diffusion models as state-of-the-art generative models and structured state space models as internal model architecture, which are particularly suited to capture long-term dependencies in time series data. We demonstrate that SSSD matches or even exceeds state-of-the-art probabilistic imputation and forecasting performance on a broad range of data sets and different missingness scenarios, including the challenging blackout-missing scenarios, where prior approaches failed to provide meaningful results.

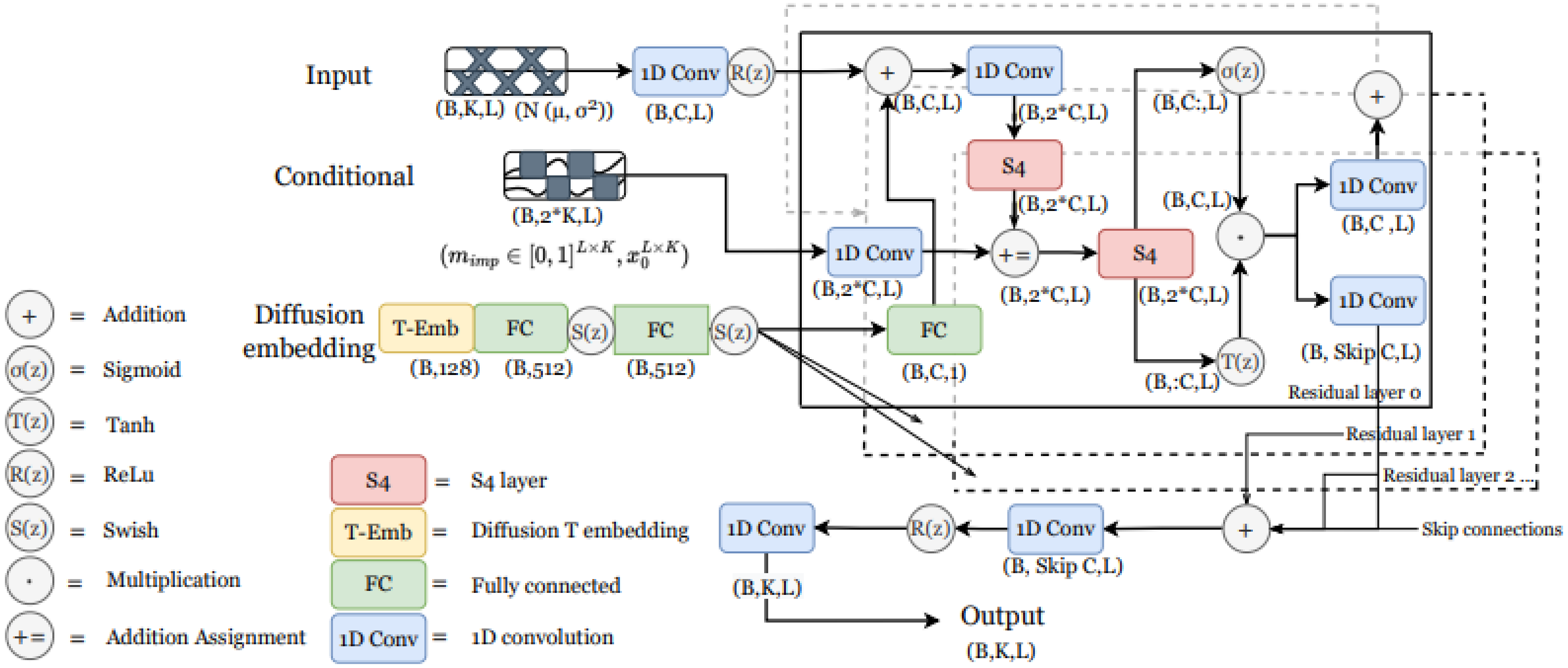
Alcaraz, J. L., & Strodthoff, N. (2022). Diffusion-based Time Series Imputation and Forecasting with Structured State Space Models. Transactions on Machine Learning Research.

Method

Diffusion-based Time Series Imputation and Forecasting with Structured State Space Models

❖ Proposed approaches

- Structured State Space models와 Diffwave를 합친 architecture 제안
- CSDI와 유사한 방식으로 결측 부분에 대해서만 Diffusion process 적용

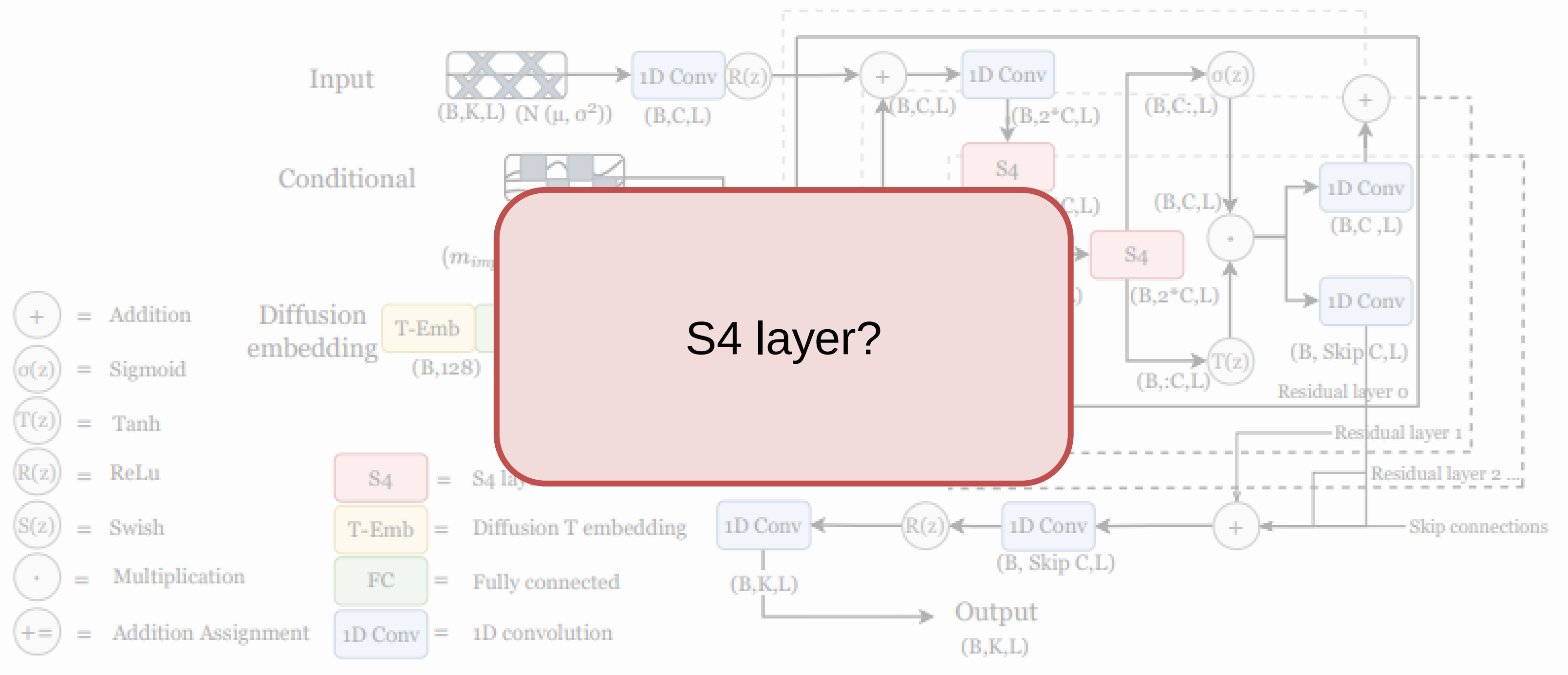


Method

Diffusion-based Time Series Imputation and Forecasting with Structured State Space Models

❖ S4 layer

- 시계열 데이터의 long-term dependencies를 다룰 수 있음



Alcaraz, J. L., & Strodthoff, N. (2022). Diffusion-based Time Series Imputation and Forecasting with Structured State Space Models. Transactions on Machine Learning Research.

Structured State Space Models

Efficiently Modeling Long Sequences with Structured State Spaces

❖ State Spaces Model

- 시간에 따라 변하는 시스템의 상태를 미분 방정식으로 표현하는 수학적 모델
- 제어 이론, 시그널 처리, 시계열 분석 같은 다양한 분야에 적용 가능

상태 방정식

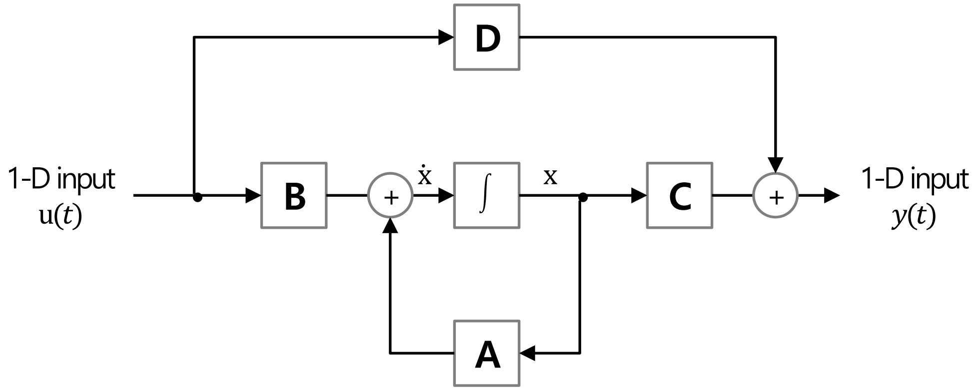
$$\mathrm{x}'(t) = A\mathrm{x}(t) + B\mathrm{u}(t)$$

$$y(t) = C\mathrm{x}(t) + D\mathrm{u}(t)$$

$\mathrm{x}(t)$: 상태 벡터
$\mathrm{x}'(t)$: 상태벡터의 시간 미분
$\mathrm{u}(t)$: 입력 벡터
$y(t)$: 출력 벡터

A: 상태 행렬
B: 입력 행렬
C: 출력 행렬
D: 피드포워드 행렬

상태 방정식 다이어그램



Gu, A., Goel, K., & Re, C. (2021, October). Efficiently Modeling Long Sequences with Structured State Spaces. In International Conference on Learning Representations.

Structured State Space Models

Efficiently Modeling Long Sequences with Structured State Spaces

❖ State Spaces Model

- 시간에 따라 변하는 시스템의 상태를 미분 방정식으로 표현하는 수학적 모델
- 제어 이론, 시그널 처리, 시계열 분석 같은 다양한 분야에 적용 가능

상태 방정식

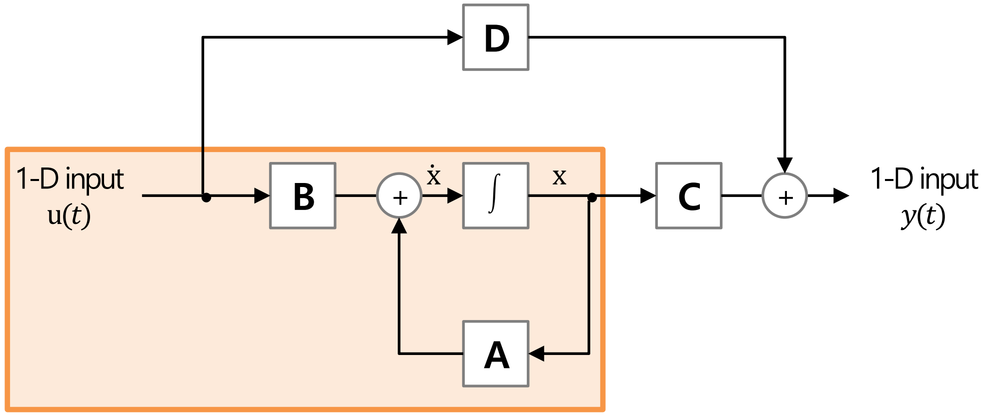
$$\boxed{x'(t) = Ax(t) + Bu(t)}$$ Input → State

$$y(t) = Cx(t) + Du(t)$$

$x(t)$: 상태 벡터
$x'(t)$: 상태벡터의 시간 미분
$u(t)$: 입력 벡터
$y(t)$: 출력 벡터

A: 상태 행렬
B: 입력 행렬
C: 출력 행렬
D: 피드포워드 행렬

상태 방정식 다이어그램



Gu, A., Goel, K., & Re, C. (2021, October). Efficiently Modeling Long Sequences with Structured State Spaces. In International Conference on Learning Representations.

Structured State Space Models

Efficiently Modeling Long Sequences with Structured State Spaces

❖ State Spaces Model

- 시간에 따라 변하는 시스템의 상태를 미분 방정식으로 표현하는 수학적 모델
- 제어 이론, 시그널 처리, 시계열 분석 같은 다양한 분야에 적용 가능

상태 방정식

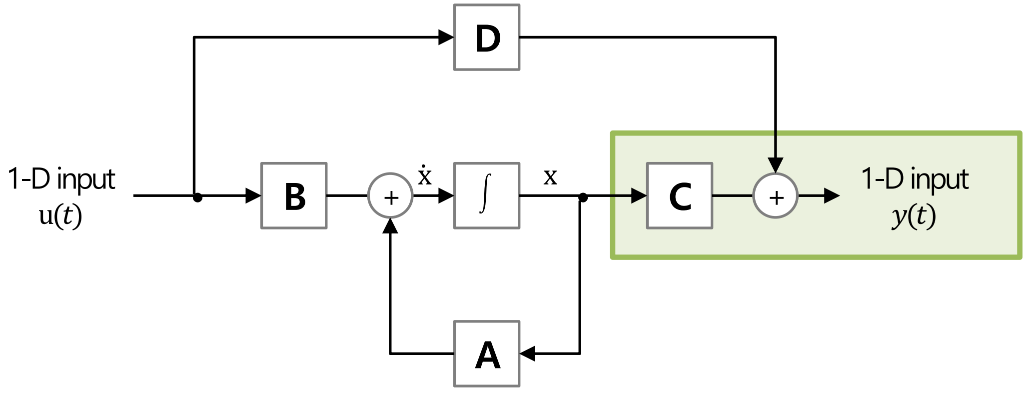
$$\mathrm{x}'(t) = A\mathrm{x}(t) + B\mathrm{u}(t)$$

$$\boxed{y(t) = C\mathrm{x}(t) + D\mathrm{u}(t)}$$ State → Output

$\mathrm{x}(t)$: 상태 벡터
$\mathrm{x}'(t)$: 상태벡터의 시간 미분
$\mathrm{u}(t)$: 입력 벡터
$y(t)$: 출력 벡터

A: 상태 행렬
B: 입력 행렬
C: 출력 행렬
D: 피드포워드 행렬

상태 방정식 다이어그램



Gu, A., Goel, K., & Re, C. (2021, October). Efficiently Modeling Long Sequences with Structured State Spaces. In International Conference on Learning Representations.

Structured State Space Models

Efficiently Modeling Long Sequences with Structured State Spaces

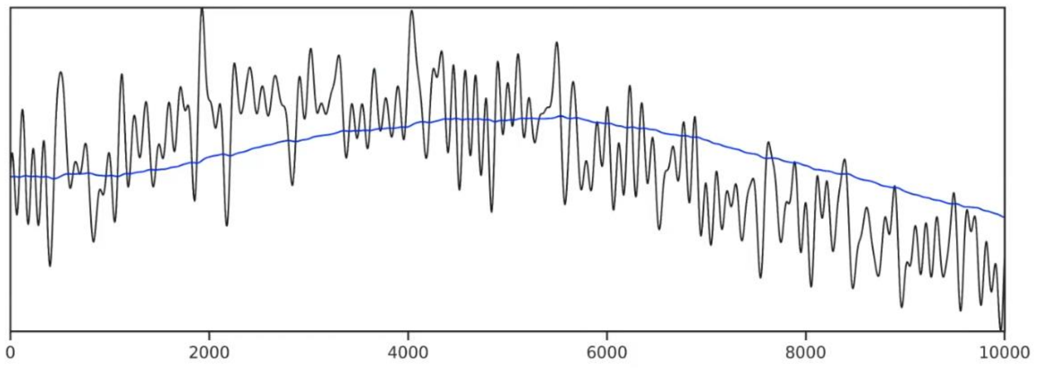
❖ SSSD Main Contribution – S4 Layer

- 시그널을 잘 표현할 수 있는 상태 벡터를 찾기

상태 방정식

$$x'(t) = \boxed{Ax(t)} + Bu(t)$$

$$y(t) = Cx(t) + Du(t)$$



Exponential Moving Average

$x(t)$: 상태 벡터
$x'(t)$: 상태벡터의 시간 미분
$u(t)$: 입력 벡터
$y(t)$: 출력 벡터

A: 상태 행렬
B: 입력 행렬
C: 출력 행렬
D: 피드포워드 행렬

Alcaraz, J. L., & Strodthoff, N. (2022). Diffusion-based Time Series Imputation and Forecasting with Structured State Space Models. Transactions on Machine Learning Research.

Structured State Space Models

Efficiently Modeling Long Sequences with Structured State Spaces

❖ SSSD Main Contribution – S4 Layer

- 시그널을 잘 표현할 수 있는 상태 벡터를 찾기

하지만 모든 상태 공간에 대한 정보를 찾기에는
Computation Cost가 너무 높음!

상태 방정식

$x'(t) = Ax(t) + Bu(t)$

$y(t) = Cx(t) + Du(t)$

x(t): 상태 벡터
x′(t): 상태벡터의 시간 미분
u(t): 입력 벡터
$y(t)$: 출력 벡터

$$\text{(HiPPO Matrix)} \qquad A_{nk} = -\begin{cases} (2n+1)^{1/2}(2k+1)^{1/2} & \text{if } n > k \\ n+1 & \text{if } n = k \\ 0 & \text{if } n < k \end{cases}$$

⇨ A 행렬을 특정 행렬로 구조화
A를 구하는 연산량 감소

Algorithm 1 S4 CONVOLUTION KERNEL (SKETCH)

Input: S4 parameters $\boldsymbol{\Lambda}, \boldsymbol{P}, \boldsymbol{Q}, \boldsymbol{B}, \boldsymbol{C} \in \mathbb{C}^N$ and step size Δ
Output: SSM convolution kernel $\overline{\boldsymbol{K}} = \mathcal{K}_L(\overline{\boldsymbol{A}}, \overline{\boldsymbol{B}}, \overline{\boldsymbol{C}})$ for $\boldsymbol{A} = \boldsymbol{\Lambda} - \boldsymbol{P}\boldsymbol{Q}^*$ (equation (5))

1: $\widetilde{\boldsymbol{C}} \leftarrow \left(\boldsymbol{I} - \overline{\boldsymbol{A}}^L\right)^* \overline{\boldsymbol{C}}$ ▷ Truncate SSM generating function (SSMGF) to length L
2: $\begin{bmatrix} k_{00}(\omega) & k_{01}(\omega) \\ k_{10}(\omega) & k_{11}(\omega) \end{bmatrix} \leftarrow \left[\widetilde{\boldsymbol{C}}\ \boldsymbol{Q}\right]^* \left(\frac{2}{\Delta}\frac{1-\omega}{1+\omega} - \boldsymbol{\Lambda}\right)^{-1} [\boldsymbol{B}\ \boldsymbol{P}]$ ▷ Black-box Cauchy kernel
3: $\hat{\boldsymbol{K}}(\omega) \leftarrow \frac{2}{1+\omega}\left[k_{00}(\omega) - k_{01}(\omega)(1+k_{11}(\omega))^{-1}k_{10}(\omega)\right]$ ▷ Woodbury Identity
4: $\hat{\boldsymbol{K}} = \{\hat{\boldsymbol{K}}(\omega) : \omega = \exp(2\pi i \frac{k}{L})\}$ ▷ Evaluate SSMGF at all roots of unity $\omega \in \Omega_L$
5: $\overline{\boldsymbol{K}} \leftarrow \mathsf{iFFT}(\hat{\boldsymbol{K}})$ ▷ Inverse Fourier Transform

$y(t)$ ↑ $x(t)$ ↑ $u(t)$

Alcaraz, J. L., & Strodthoff, N. (2022). Diffusion-based Time Series Imputation and Forecasting with Structured State Space Models. Transactions on Machine Learning Research.

SSSD

Diffusion-based Time Series Imputation and Forecasting with Structured State Space Models

❖ SSSD Main Contribution – S4 Layer

- Long Sequence 데이터에 특화 되어있고, Time Series Data에 적합

Model	ListOps	Text	Retrieval	Image	Pathfinder	Path-X	Avg
Random	10.00	50.00	50.00	10.00	50.00	50.00	36.67
Transformer	36.37	64.27	57.46	42.44	71.40	✗	53.66
Local Attention	15.82	52.98	53.39	41.46	66.63	✗	46.71
Sparse Trans.	17.07	63.58	59.59	44.24	71.71	✗	51.03
Longformer	35.63	62.85	56.89	42.22	69.71	✗	52.88
Linformer	35.70	53.94	52.27	38.56	76.34	✗	51.14
Reformer	37.27	56.10	53.40	38.07	68.50	✗	50.56
Sinkhorn Trans.	33.67	61.20	53.83	41.23	67.45	✗	51.23
Synthesizer	36.99	61.68	54.67	41.61	69.45	✗	52.40
BigBird	36.05	64.02	59.29	40.83	74.87	✗	54.17
Linear Trans.	16.13	65.90	53.09	42.34	75.30	✗	50.46
Performer	18.01	65.40	53.82	42.77	77.05	✗	51.18
FNet	35.33	65.11	59.61	38.67	77.80	✗	54.42
Nyströmformer	37.15	65.52	79.56	41.58	70.94	✗	57.46
Luna-256	37.25	64.57	79.29	47.38	77.72	✗	59.37
S4	**58.35**	**76.02**	**87.09**	**87.26**	**86.05**	**88.10**	**80.48**

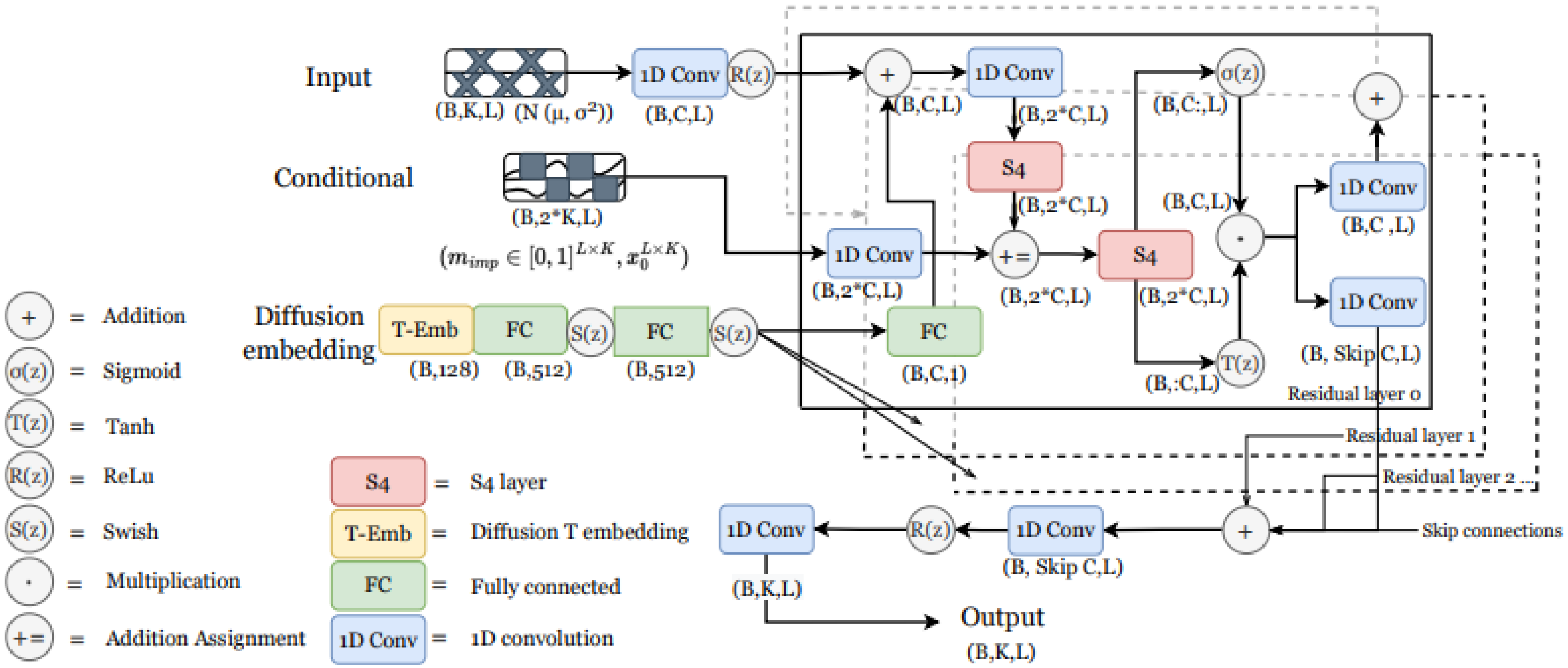
Alcaraz, J. L., & Strodthoff, N. (2022). Diffusion-based Time Series Imputation and Forecasting with Structured State Space Models. Transactions on Machine Learning Research.

Method

Diffusion-based Time Series Imputation and Forecasting with Structured State Space Models

❖ Diffusion Process

- Sequential Data의 특징을 반영하기 더욱 수월한 S4 layer을 사용함으로써 Time Series Imputation에서 좋은 성능을 냄



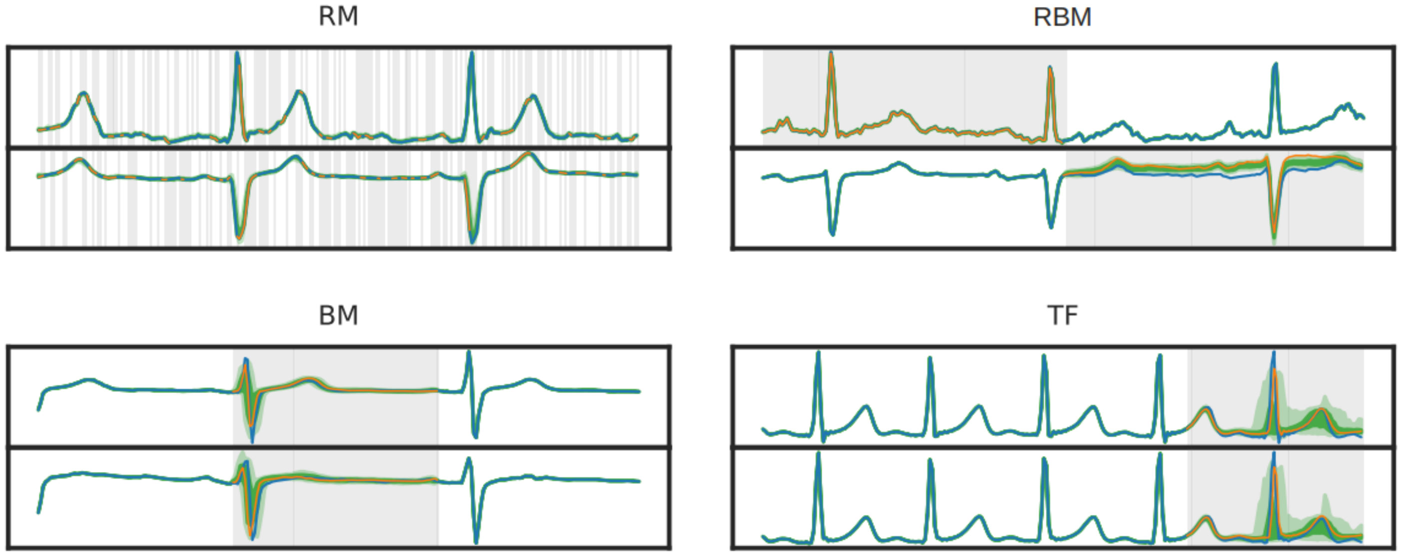
Alcaraz, J. L., & Strodthoff, N. (2022). Diffusion-based Time Series Imputation and Forecasting with Structured State Space Models. Transactions on Machine Learning Research.

Experiment

Diffusion-based Time Series Imputation and Forecasting with Structured State Space Models

❖ Missingness Scenarios 정의

- **R**andom **M**issing : Random 위치에서 결측치가 발생하는 상황
- **R**andom **B**lock **M**issing : Random 위치에서 결측치가 발생하는 상황
- **B**lackout **M**issing : 특정한 위치에서 결측치가 발생하는 상황
- **T**ime series **F**orecasting : 시계열 예측, 시계열의 마지막 부분에서 결측치가 생기는 Special 형태로 봐도 무방



Alcaraz, J. L., & Strodthoff, N. (2022). Diffusion-based Time Series Imputation and Forecasting with Structured State Space Models. Transactions on Machine Learning Research.

Experiment

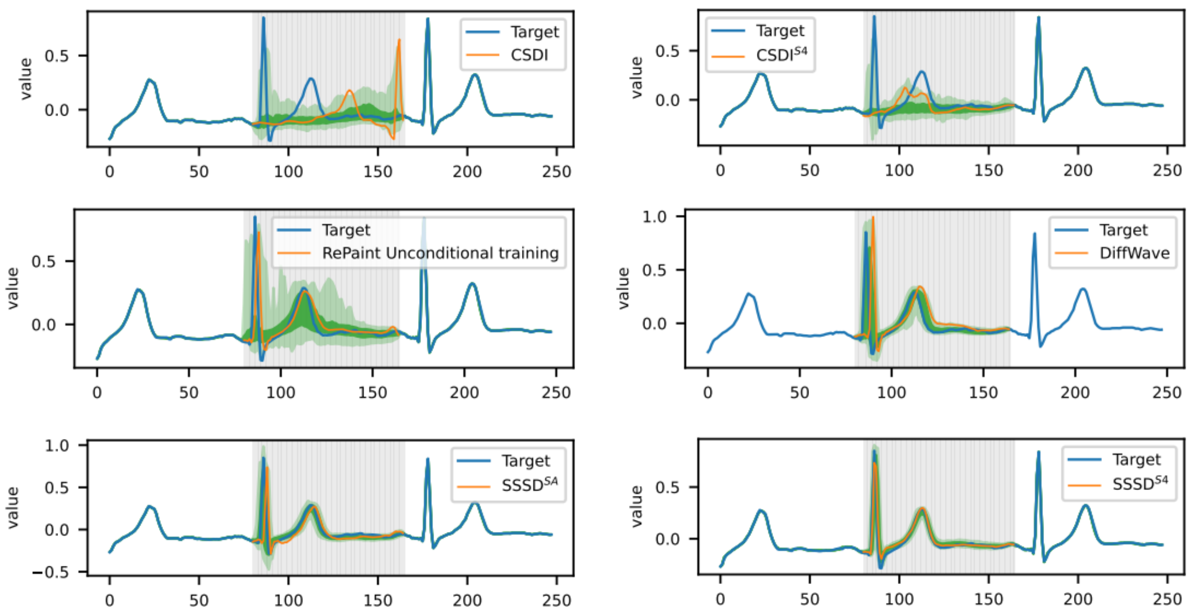
Diffusion-based Time Series Imputation and Forecasting with Structured State Space Models

❖ ECG PTB-XL dataset

- Sampling rate 100Hz & 250 Time steps
- 평가 지표 : Test set에서 각 샘플에 대해 생성된 100개 샘플의 MAE, RMSE 평균

Model	MAE	RMSE
20% RM on PTB-XL		
LAMC	0.0678	0.1309
CSDI	0.0038±2e-6	0.0189±5e-5
DiffWave	0.0043±4e-4	0.0177±4e-4
CSDI^{S4}	**0.0031±1e-7**	0.0171±6e-4
SSSD^{SA}	0.0045±3e-7	0.0181±4e-6
SSSD^{S4}	0.0034±4e-6	**0.0119±1e-4**
20% RBM on PTB-XL		
LAMC	0.0759	0.1498
CSDI	0.0186±1e-5	0.0435±2e-4
DiffWave	0.0250±1e-3	0.0808±5e-3
CSDI^{S4}	0.0222±2e-5	0.0573±1e-3
SSSD^{SA}	0.0170±1e-4	0.0492±1e-2
SSSD^{S4}	**0.0103±3e-3**	**0.0226±9e-4**
20% BM on PTB-XL		
LAMC	0.0840	0.1171
CSDI	0.1054±4e-5	0.2254±7e-5
DiffWave	0.0451±7e-4	0.1378±5e-3
CSDI^{S4}	0.0792±2e-4	0.1879±1e-4
SSSD^{SA}	0.0435±3e-3	0.1167±1e-2
SSSD^{S4}	**0.0324±3e-3**	**0.0832±8e-3**

Black out Missing에 대한 결과



Alcaraz, J. L., & Strodthoff, N. (2022). Diffusion-based Time Series Imputation and Forecasting with Structured State Space Models. Transactions on Machine Learning Research.

Experiment

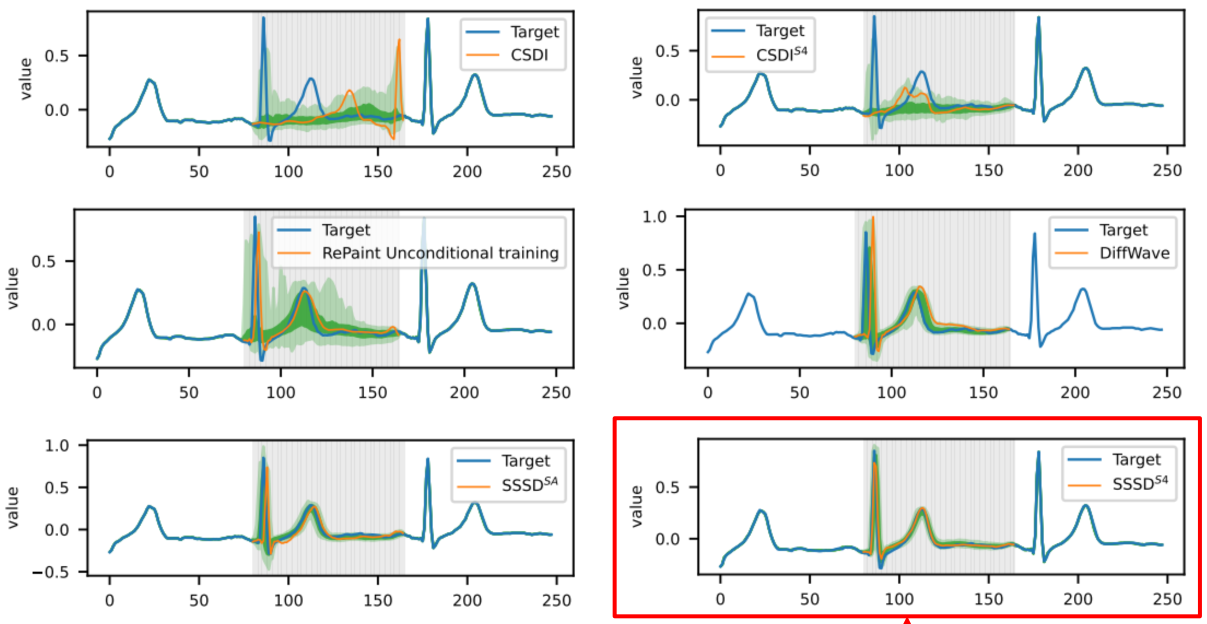
Diffusion-based Time Series Imputation and Forecasting with Structured State Space Models

❖ ECG PTB-XL dataset

- Sampling rate 100Hz & 250 Time steps
- 평가 지표 : Test set에서 각 샘플에 대해 생성된 100개 샘플의 MAE, RMSE 평균

Model	MAE	RMSE
20% RM on PTB-XL		
LAMC	0.0678	0.1309
CSDI	0.0038±2e-6	0.0189±5e-5
DiffWave	0.0043±4e-4	0.0177±4e-4
CSDI^{S4}	**0.0031±1e-7**	0.0171±6e-4
SSSD^{SA}	0.0045±3e-7	0.0181±4e-6
SSSD^{S4}	0.0034±4e-6	**0.0119±1e-4**
20% RBM on PTB-XL		
LAMC	0.0759	0.1498
CSDI	0.0186±1e-5	0.0435±2e-4
DiffWave	0.0250±1e-3	0.0808±5e-3
CSDI^{S4}	0.0222±2e-5	0.0573±1e-3
SSSD^{SA}	0.0170±1e-4	0.0492±1e-2
SSSD^{S4}	**0.0103±3e-3**	**0.0226±9e-4**
20% BM on PTB-XL		
LAMC	0.0840	0.1171
CSDI	0.1054±4e-5	0.2254±7e-5
DiffWave	0.0451±7e-4	0.1378±5e-3
CSDI^{S4}	0.0792±2e-4	0.1879±1e-4
SSSD^{SA}	0.0435±3e-3	0.1167±1e-2
SSSD^{S4}	**0.0324±3e-3**	**0.0832±8e-3**

Black out Missing에 대한 결과



Alcaraz, J. L., & Strodthoff, N. (2022). Diffusion-based Time Series Imputation and Forecasting with Structured State Space Models. Transactions on Machine Learning Research.

Experiment

Diffusion-based Time Series Imputation and Forecasting with Structured State Space Models

❖ Electricity dataset

- High dimensional (100 Channel 이상, 370 Features)
- 준수한 성능을 보여줌
- 평가 지표 : 각 샘플에 대해 3번 생성한 샘플의 MAE, RMSE, MRE 평균

Model	MAE	RMSE	MRE
10% RM on Electricity			
Median	2.056	2.732	110%
M-RNN	1.244	1.867	66.6%
GP-VAE	1.094	1.565	58.6%
BRITS	0.847	1.322	45.3%
Transformer	0.823	1.301	44.0%
SAITS	0.735	1.162	39.4%
CSDI	1.510±3e-3	15.012±4e-2	81.10%±1e-3
SSSD^{S4}	**0.345±1e-4**	**0.554±5e-5**	**18.4%±5e-5**
30% RM on Electricity			
Median	2.055	2.732	110%
M-RNN	1.258	1.876	67.3%
GP-VAE	1.057	1.571	56.6%
BRITS	0.943	1.435	50.4%
Transformer	0.846	1.321	45.3%
SAITS	0.790	1.223	42.3%
CSDI	0.921±8e-3	8.372±7e-2	49.27%±4e-3
SSSD^{S4}	**0.407±5e-4**	**0.625±1e-4**	**21.8±0%**
50% RM on Electricity			
Median	2.053	2.728	109%
M-RNN	1.283	1.902	68.7%
GP-VAE	1.097	1.572	58.8%
BRITS	1.037	1.538	55.5%
Transformer	0.895	1.410	47.9%
SAITS	0.876	1.377	46.9%
CSDI	**0.278±4e-3**	2.371±3e-2	**14.93%±1e-3**
SSSD^{S4}	0.532±1e-4	**0.821±1e-4**	28.5%±1e-4

Alcaraz, J. L., & Strodthoff, N. (2022). Diffusion-based Time Series Imputation and Forecasting with Structured State Space Models. Transactions on Machine Learning Research.

Summary

Diffusion-based Time Series Imputation and Forecasting with Structured State Space Models

❖ SSSD

- Structured State Space model과 Diffusion model을 이용한 Time Series Imputation model
- 효율적으로 장기의존성을 포착하여 결측치 대체 성능이 우수함

❖ Limitation

- 샘플링 속도가 느림
- 입력 채널의 수가 100개 이상인 경우 모델이 수렴하지 못하는 경우가 있음

Rasul, K., Seward, C., Schuster, I., & Vollgraf, R. (2021, July). Autoregressive denoising diffusion models for multivariate probabilistic time series forecasting. In International Conference on Machine Learning (pp. 8857-8868). PMLR.

Conclusion

Diffusion Models for Time Series

❖ Diffusion Models

- 높은 안정성, 높은 품질의 생산성, 복잡한 분포 학습 가능

❖ Time Series Forecasting

- TimeGrad: 다변량 시계열 예측을 위한 Diffusion Model

❖ Time Series Imputation

- CSDI: Masked Language Model에 영감을 받은 Time Series Imputation Diffusion Model → 2D Transformer Layer 사용
- SSSD: CSDI에 장기의존성 문제를 해결하기 위해 Structured State Space Model을 도입한 Time Series Imputation Diffusion Model

고맙습니다