随机扩散：一种基于扩散的随机时间序列预测模型

汇报人: 韦浩文

时间：2025-8-10

目录

1 背景

背景

扩散模型在生成式时间序列预测中得到了广泛应用。然而，现有的基于扩散的时间序列预测模型，通常将时间序列建模和扩散生成视为相对独立的步骤，并且扩散模型的潜变量往往从单峰分布中采样。这种设计在处理异质时间序列数据时存在局限性。

StochDiff

1.将扩散过程整合到时序建模的每个时间步中

2.设计了一种逐步数据驱动的先验

首个将扩散过程整合到时间序列建模阶段，并设计了逐步数据驱动先验的、基于扩散的模型。

2

方法

DDPM

正向扩散　　人为设计的加噪过程

正向扩散第n步的先验分布（正向扩散第n步的加噪规则）

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

正向扩散联合分布

$$q(\boldsymbol{x}^{1:N}|\boldsymbol{x}^0) := \prod_{n=1}^{N} q(\boldsymbol{x}^n|\boldsymbol{x}^{n-1})$$

正向扩散从 $\boldsymbol{x}^0$ 直接采样 $\boldsymbol{x}^n$ 的闭式解

$$q(\boldsymbol{x}^n|\boldsymbol{x}^0) = \mathcal{N}(\boldsymbol{x}^n; \sqrt{1-\overline{\alpha}_n}\boldsymbol{x}^0, (1-\overline{\alpha}_n)\boldsymbol{I})$$

$$\alpha^n := 1-\beta_n \quad \overline{\alpha}_n := \prod_{i=1}^{n} \alpha_i$$

DDPM

反向扩散　　模型学习的去噪过程

反向扩散第n步的后验分布（反向扩散第n步的去噪规则）

$$p_\theta(\boldsymbol{x}^{n-1}|\boldsymbol{x}^n) := \mathcal{N}(\boldsymbol{x}^{n-1}; \boxed{\boldsymbol{\mu}_\theta(\boldsymbol{x}^n, n)}, \boxed{\boldsymbol{\delta}_\theta(\boldsymbol{x}^n, n)}) \longrightarrow \boldsymbol{\sigma}_n^2 \boldsymbol{I}$$

$$\boldsymbol{\sigma}_n^2 = \frac{1-\overline{\alpha}_{n-1}}{1-\overline{\alpha}_n}\beta_n$$

反向扩散联合分布

$$p_\theta(\boldsymbol{x}^0, \boldsymbol{x}^1, \ldots, \boldsymbol{x}^N) = p_\theta(\boldsymbol{x}^{0:N}) := p(\boldsymbol{x}^N)\prod_{n=1}^{N} p_\theta(\boldsymbol{x}^{n-1}|\boldsymbol{x}^n)$$

$$p_\theta(x^0) := \int p_\theta(x^{0:N})dx^{1:N}$$

DDPM

训练

$$\mathcal{L}_n = D_{KL}(q(\boldsymbol{x}^{n-1}|\boldsymbol{x}^n)||p_\theta(\boldsymbol{x}^{n-1}|\boldsymbol{x}^n))$$

$$q(\boldsymbol{x}^{n-1}|\boldsymbol{x}^n, \boldsymbol{x}^0) = \mathcal{N}(\boldsymbol{x}^{n-1}; \tilde{\boldsymbol{\mu}}_n(\boldsymbol{x}^n, \boldsymbol{x}^0), \tilde{\beta}_n \boldsymbol{I})$$

$$\tilde{\boldsymbol{\mu}}_n(\boldsymbol{x}^n, \boldsymbol{x}^0) := \frac{\sqrt{\overline{\alpha}_n - 1}\beta_n}{1-\overline{\alpha}_n}\boldsymbol{x}^0 + \frac{\sqrt{\alpha_n}(1-\overline{\alpha}_{n-1})}{1-\overline{\alpha}_n}\boldsymbol{x}^n \qquad \tilde{\beta}_n := \frac{1-\overline{\alpha}_{n-1}}{1-\overline{\alpha}_n}\beta_n$$

$$\mathcal{L}_n = \frac{1}{2\boldsymbol{\sigma}_n^2}||\tilde{\boldsymbol{\mu}}_n(\boldsymbol{x}^n, \boldsymbol{x}^0, n) - \boldsymbol{\mu}_\theta(\boldsymbol{x}^n, n)||^2$$

DDPM

均值函数的建模

$$\boldsymbol{\epsilon}_\theta(\boldsymbol{x}^n, n) \longrightarrow \boldsymbol{\mu}(\boldsymbol{\epsilon}_\theta)$$

$$\boldsymbol{x}_\theta(\boldsymbol{x}^n, n) \longrightarrow \boldsymbol{\mu}(\boldsymbol{x}_\theta)$$

$$\boldsymbol{\mu}(\boldsymbol{x}_\theta) = \frac{\sqrt{\alpha_n}(1-\overline{\alpha}_{n-1})}{1-\overline{\alpha}_n}\boldsymbol{x}^n + \frac{\sqrt{\overline{\alpha}_{n-1}}\beta_n}{1-\overline{\alpha}_k}\boldsymbol{x}_\theta(\boldsymbol{x}^n, n)$$

$$\mathcal{L}_{\boldsymbol{x}} = \mathbb{E}_{n,\boldsymbol{x}^0}[||\boldsymbol{x}^0 - \boldsymbol{x}_\theta(\boldsymbol{x}^n, n)||^2]$$

概率性时间序列预测

$$x_{1:T}^0 = \{x_1^0, x_2^0, \ldots, x_{t_0}^0, \ldots, x_T^0\}$$

$$q_{\mathcal{X}}(\boldsymbol{x}_{t_0:T}^0 | \boldsymbol{x}_{1:t_0-1}^0) = \prod_{t=t_0}^{T} q_{\mathcal{X}}(\boldsymbol{x}_t^0 | \boxed{\boldsymbol{x}_{1:t-1}^0})$$

StochDiff

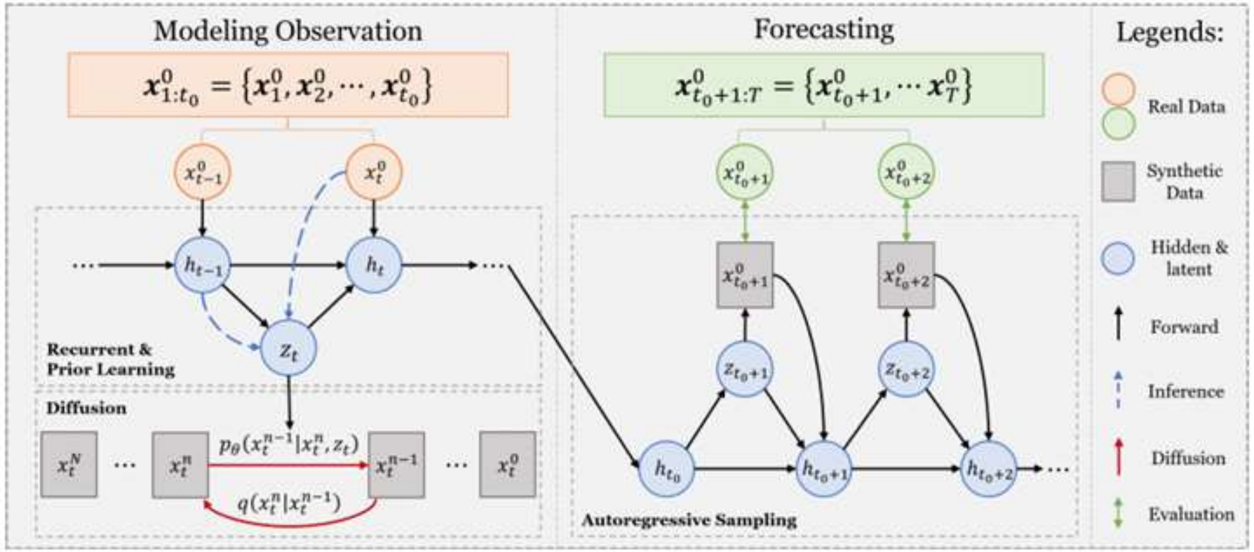


Figure 1: Stochastic Diffusion for Time Series Forecasting

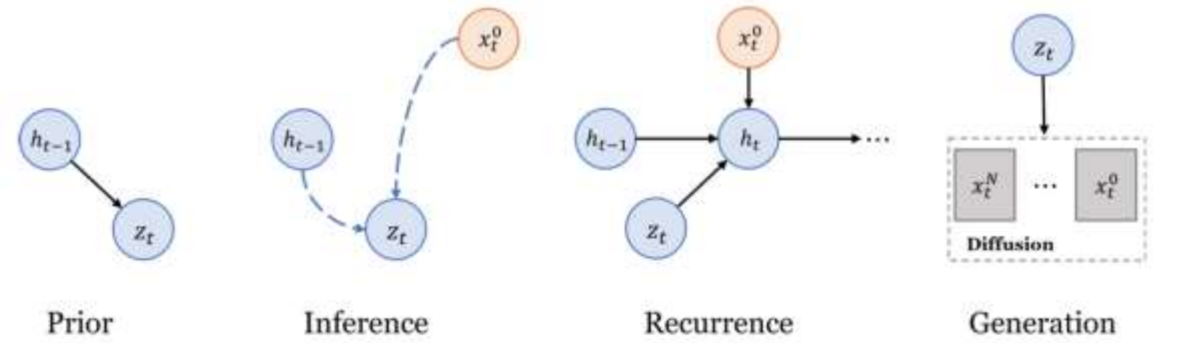
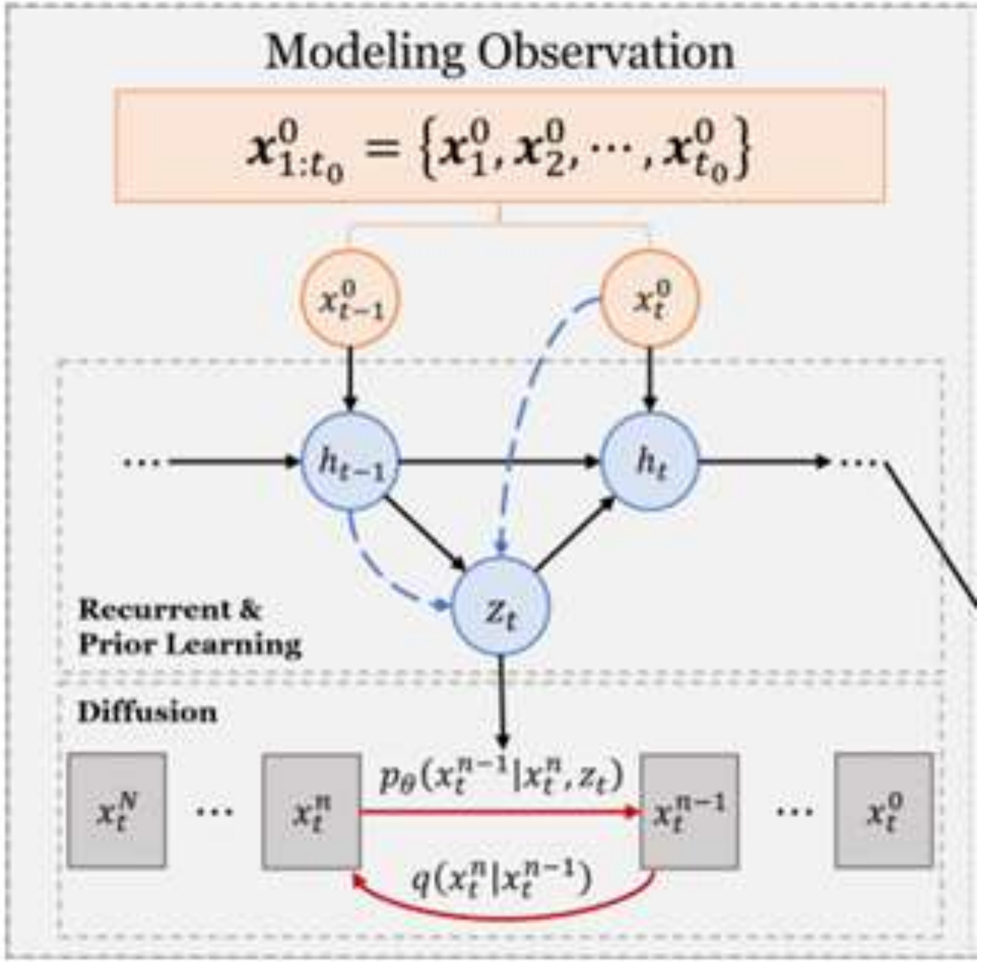


Figure 2: Graphic illustration of the modeling operations of StochDiff. (1) Obtaining conditional prior. (2) Inference of approximate posterior. (3) Hidden state update via sequential model. (4) Data generation via diffusion model.

1.将扩散过程整合到时序建模的每个时间步中

2.设计了一种逐步数据驱动的先验

StochDiff-Modeling



数据驱动先验

$$z_t \sim p_z(z_t|x^0_{1:t-1}, z_{1:t-1}) := \mathcal{N}(\boxed{\hat{\mu}_\theta(h_{t-1})}, \boxed{\hat{\delta}_\theta(h_{t-1})})$$

$$h_{t-1} = f_\theta(x^0_{t-1}, z_{t-1}, h_{t-2})$$

真实后验

$$p(z_t|x^0_{1:t}, z_{1:t-1})$$

近似后验

$$q_z(z_t|x^0_{1:t}, z_{1:t-1}) := \mathcal{N}(\mu_{z,t}(h_{t-1}, x^0_t), \delta_{z,t}(h_{t-1}, x^0_t))$$

将先验变量作为反向扩散过程的条件 $\sigma^2_n I$

$$p_\theta(x^{n-1}_t|x^n_t, z_t) = \mathcal{N}(x^{n-1}_t; \mu_C(x^n_t, n, z_t), \delta_C(x^n_t, n, z_t))$$

StochDiff-Forecasting

未来时间序列的联合分布

$$p_\theta(\boldsymbol{x}^0_{t_0:T}|\boldsymbol{x}^0_{1:t_0-1}, \boldsymbol{z}_{1:t_0-1}) =$$

$$\int_{\boldsymbol{x}^{1:N}_{t_0:T}} \int_{\boldsymbol{z}_{t_0:T}} \prod_{t=t_0}^{T} p(\boldsymbol{x}^N_t|\boldsymbol{z}_t) \prod_{n=1}^{N} p_\theta(\boldsymbol{x}^{n-1}_t|\boldsymbol{x}^n_t, \boldsymbol{z}_t) p_z(\boldsymbol{z}_t|\boldsymbol{x}^0_{1:t-1}, \boldsymbol{z}_{1:t-1})$$

$$d\boldsymbol{z}_{t_0:T} d\boldsymbol{x}^{1:N}_{t_0:T}.$$

反向扩散每步去噪

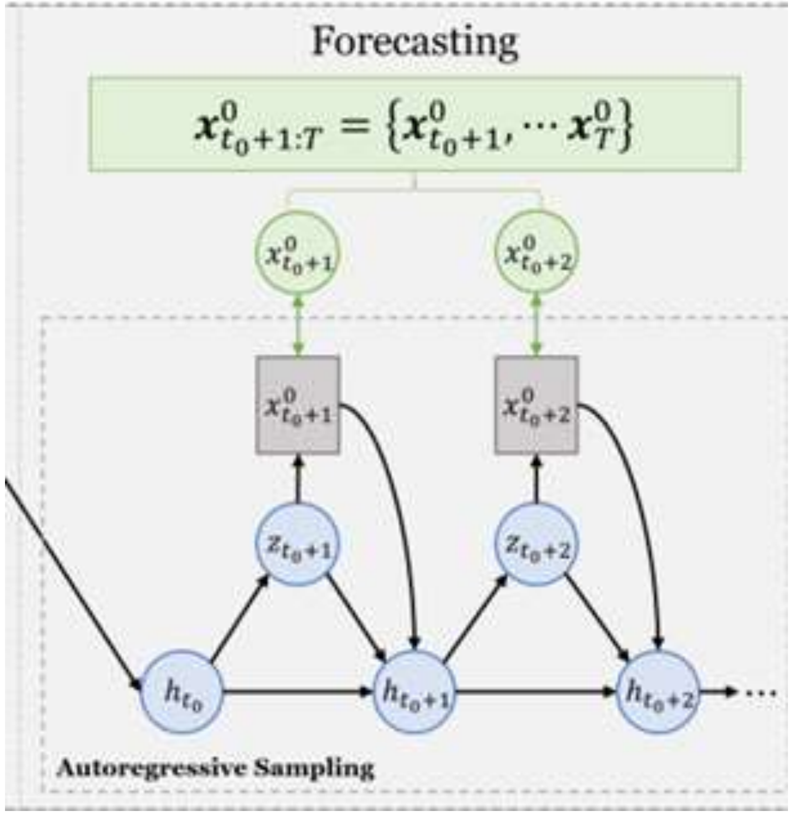
$$\boldsymbol{x}^{n-1}_t = \frac{\sqrt{\alpha_n}(1-\overline{\alpha}_{n-1})}{1-\overline{\alpha}_n}\boldsymbol{x}^n_t + \frac{\sqrt{\overline{\alpha}_{n-1}}\beta_n}{1-\overline{\alpha}_k}\boldsymbol{x}_\theta(\boldsymbol{x}^n_t, n, \boldsymbol{z}_t)$$

StochDiff训练的目标

$$\mathcal{L}_{dual} = \sum_{t=1} D_{KL}(q_z(\boldsymbol{z}_t|\boldsymbol{x}^0_{1:t}, \boldsymbol{z}_{1:t-1})||p_z(\boldsymbol{z}_t|\boldsymbol{x}^0_{1:t-1}, \boldsymbol{z}_{1:t-1}))$$

$$+ \mathbb{E}_{\boldsymbol{x}^0_t, n, \boldsymbol{z}_t}[||\boldsymbol{x}^0_t - \boldsymbol{x}_\theta(\boldsymbol{x}^n_t, n, \boldsymbol{z}_t)||^2].$$

StochDiff-Forecasting



Algorithm 1 Training via Time Series modeling

Input Training time series data $\boldsymbol{x}_{1:t_0}$.
Initialize $h_0 = 0, \mathcal{L}_{total} = 0$.
repeat
  for $t = 1$ to t_0 **do**
    Sample $n \sim \mathcal{U}(\{1, 2, \dots, N\})$.
    Sample $\epsilon \sim \mathcal{N}(0, \boldsymbol{I})$.
    Observe $\boldsymbol{x}_t$ as $\boldsymbol{x}_t^0$.
    Obtain p_z, q_z, and $\boldsymbol{z}_t \sim q_z$ from equation 10, 11.
    Obtain reconstructed $\boldsymbol{x}_t^0$ based on $\boldsymbol{z}_t$ and equation equation 14.
    Update h_t via $\boldsymbol{f}_\theta(\boldsymbol{x}_t, \boldsymbol{z}_t, h_{t-1})$.
    Calculate loss function $\mathcal{L}_{dual}$ in equation 16.
    $\mathcal{L}_{total} += \mathcal{L}_{dual}$.
  end for
  Take gradient descent step on $\nabla \mathcal{L}_{total}$, and update model parameters.
until converged

Algorithm 2 Autoregressive Forecasting

Require: trained denoising network $\boldsymbol{x}_\theta$, recurrent network $\boldsymbol{f}_\theta$.
Input Test time series data $\boldsymbol{x}_{1:T}$.
Initialize $h_0 = 0$.
for $t = 1$ to t_0 **do**
  Observe $\boldsymbol{x}_t$ as $\boldsymbol{x}_t^0$.
  Obtain $\boldsymbol{z}_t \sim q_z$ from equation 11.
  Update h_t via $\boldsymbol{f}_\theta(\boldsymbol{x}_t, \boldsymbol{z}_t, h_{t-1})$.
end for
Obtain h_{t_0} (end point of the previous **for** loop).
for $t = t_0 + 1$ to T **do**
  Obtain $\boldsymbol{z}_t \sim p_z$ from equation 10.
  for $n = N$ to 1 **do**
    Sample $\hat{\boldsymbol{x}}_t^{n-1}$ using equation 14.
  end for
  Update h_t via $\boldsymbol{f}_\theta(\hat{\boldsymbol{x}}_t^0, \boldsymbol{z}_t, h_{t-1})$.
  Obtain $\hat{\boldsymbol{x}}_t^0$ for $\hat{\boldsymbol{x}}_{t_0+1:T}$.
end for
return $\hat{\boldsymbol{x}}_{t_0+1:T}$

高斯混合模型GMM

3 实验

实验-数据集、评估指标

数据集

Table 1: Statistical details of the datasets. Values in parenthesis represents the number of sampled data points.

Datasets	Window Size	Forecasting	Dimension	Seasonality	Stationarity
Exchange	3 months (90)	1 week (7)	22	0.316	0.567
Weather	1 day (144)	1 day (144)	21	0.649	9E-14
Electricity	7 days (168)	1 day (24)	321	0.923	2E-24
Solar	1 days (288)	1/2 day (144)	56	0.891	9E-15
ECochG	1/2 minute (50)	7 seconds (10)	218	0.357	0.153
MMG	3 seconds (100)	0.3 second (10)	148	0.179	6E-27

评估指标

- 1.归一化均方根误差NRMSE

$$\mathrm{NRMSE} := \frac{\sqrt{\mathbb{E}((x-\hat{x})^2)}}{\mathbb{E}(x)}$$

- 2. CRPS

$$\mathrm{CRPS}(F, x) := \int_{\mathbb{R}} (F(y) - \mathbb{I}(x \leq y))^2 dy$$

$$\mathrm{CRPSsum} = \mathbb{E}_t \left[\mathrm{CRPS} \left(\hat{F}_{\mathrm{sum}}(y), \sum_{i=1}^{d} x_t^{(i)} \right) \right]$$

$$\hat{F}_{\mathrm{sum}}(y) = \frac{1}{S} \sum_{s=1}^{S} \mathbb{I}(x^s \leq y)$$

实验-StochDiff网络细节

序列模型-LSTM
先验变量-VAE
扩散模型-正向扩散（主要涉及数学计算）、反向扩散（Attention-Net）

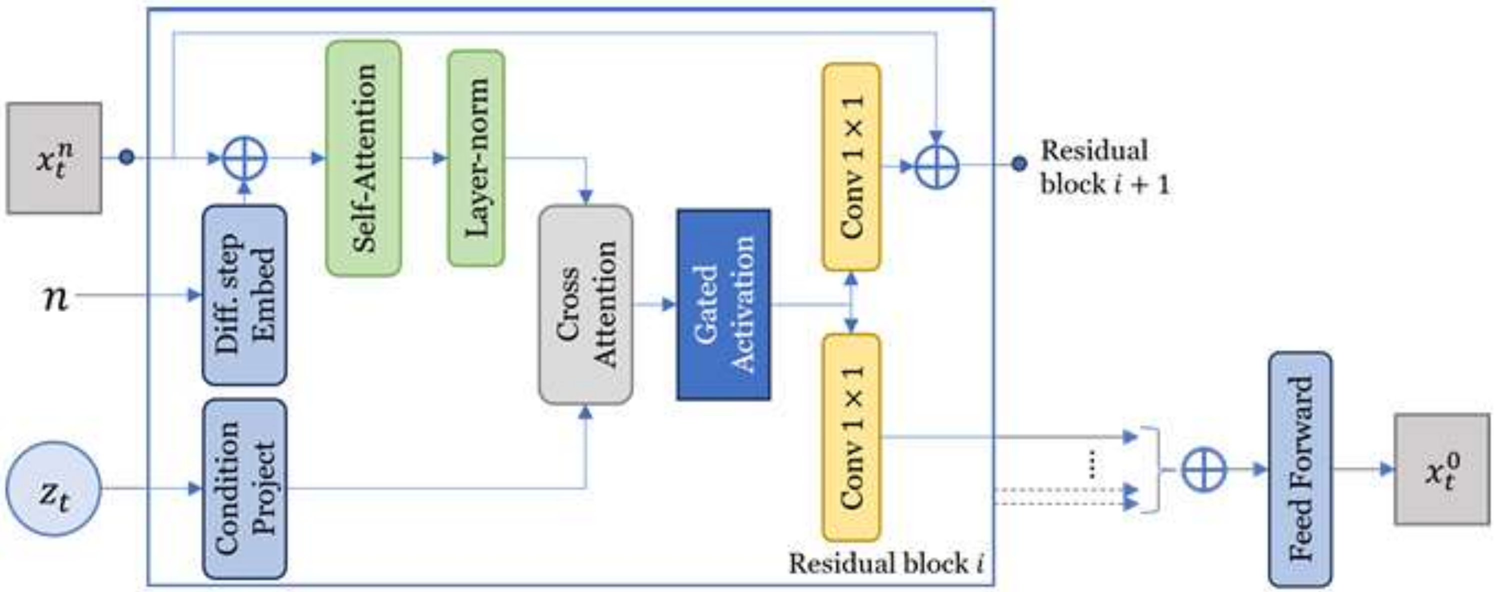


Figure 3: Attention-Net

实验-对比实验

Table 2: $NRMSE$ results. Lower values are better.

Model	Exchange	Weather	Electricity	Solar	ECochG	MMG
Transformer MAF	0.075±0.007	0.812±0.019	0.052±0.005	0.917±0.056	1.877±0.043	3.067±0.104
TimeGrad	0.066±0.007	0.691±0.017	0.043±0.003	0.973±0.056	1.652±0.025	2.977±0.101
SSSD	0.065±0.006	0.755±0.041	0.094±0.005	1.058±0.053	0.965±0.041	2.184±0.105
TimeDiff	0.074±0.002	0.702±0.025	0.045±0.003	0.882±0.044	0.916±0.033	1.565±0.091
TMDM	0.063±0.005	0.564±0.019	0.063±0.004	0.829±0.038	1.074±0.052	5.366±0.114
MG-TSD	0.075±0.005	0.696±0.023	**0.032±0.002**	**0.756±0.051**	0.994±0.045	1.716±0.087
StochDiff (ours)	**0.052±0.007**	**0.491±0.014**	0.048±0.002	0.812±0.042	**0.859±0.031**	**1.529±0.094**

Table 3: $CRPS_{sum}$ results. Lower values are better.

Model	Exchange	Weather	Electricity	Solar	ECochG	MMG
Transformer MAF	0.009±0.004	0.546±0.057	0.079±0.003	0.476±0.067	1.186±0.011	1.235±0.059
TimeGrad	0.010±0.002	0.527±0.015	0.024±0.002	0.416±0.036	1.021±0.002	1.057±0.041
SSSD	0.010±0.001	0.701±0.047	0.065±0.005	0.506±0.053	0.781±0.003	1.372±0.025
TimeDiff	0.012±0.002	0.635±0.019	0.036±0.004	0.423±0.071	0.516±0.002	0.948±0.020
TMDM	0.010±0.001	0.618±0.029	0.042±0.005	0.382±0.039	0.703±0.004	1.495±0.043
MG-TSD	0.009±0.000	0.596±0.027	**0.019±0.001**	**0.375±0.067**	0.683±0.002	0.961±0.031
StochDiff (ours)	**0.008±0.001**	**0.521±0.012**	0.029±0.003	0.391±0.052	**0.435±0.003**	**0.818±0.023**

➢ **基线方法**

扩散模型（经典模型：TimeGrad、SSSD、TimeDiff + 最新模型：TMDM、MG-TSD）

非扩散模型（Transformer MAF）

实验-消融实验

Table 4: Ablation Study on *ECochG* and *MMG* datasets

Model	ECochG		MMG	
	NRMSE	CRPS	NRMSE	CRPS
LSTM	1.003±0.042	-	3.657±0.138	-
vLSTM-$\mathcal{N}(0,1)$	0.982±0.036	0.623±0.014	3.315±0.114	1.767±0.041
vLSTM-diffusion	0.956±0.035	0.505±0.017	3.001±0.112	1.639±0.032
StochDiff (ours)	**0.859±0.031**	**0.435±0.003**	**1.529±0.094**	**0.818±0.023**

➢ **分析**

LSTM——vLSTM-N(0,1)——vLSTM-diffusion——StochDiff

1-2 VAE

2-3 扩散模型

3-4 逐步数据驱动的先验

4 总结

总结

- **总结:** 本文提出了一种新型的扩散模型，用于高随机性的时间序列预测。它通过设计一种逐步的、数据驱动的先验，将扩散过程整合到时序建模的每个时间步，能够更好的建模高随机性时间序列数据中的动态性和不确定性。
- **展望**：未来可以探索直接把先验变量融入潜变量的方法、组合多种序列模型等设计来改进现有的研究。

谢谢大家!

汇报人: 韦浩文

时间：2025-8-10