Autoformer and Autoregressive Denoising Diffusion Models for Time Series Forecasting

Гущин Никита

ШАД

МГУ им. М.В. Ломоносова

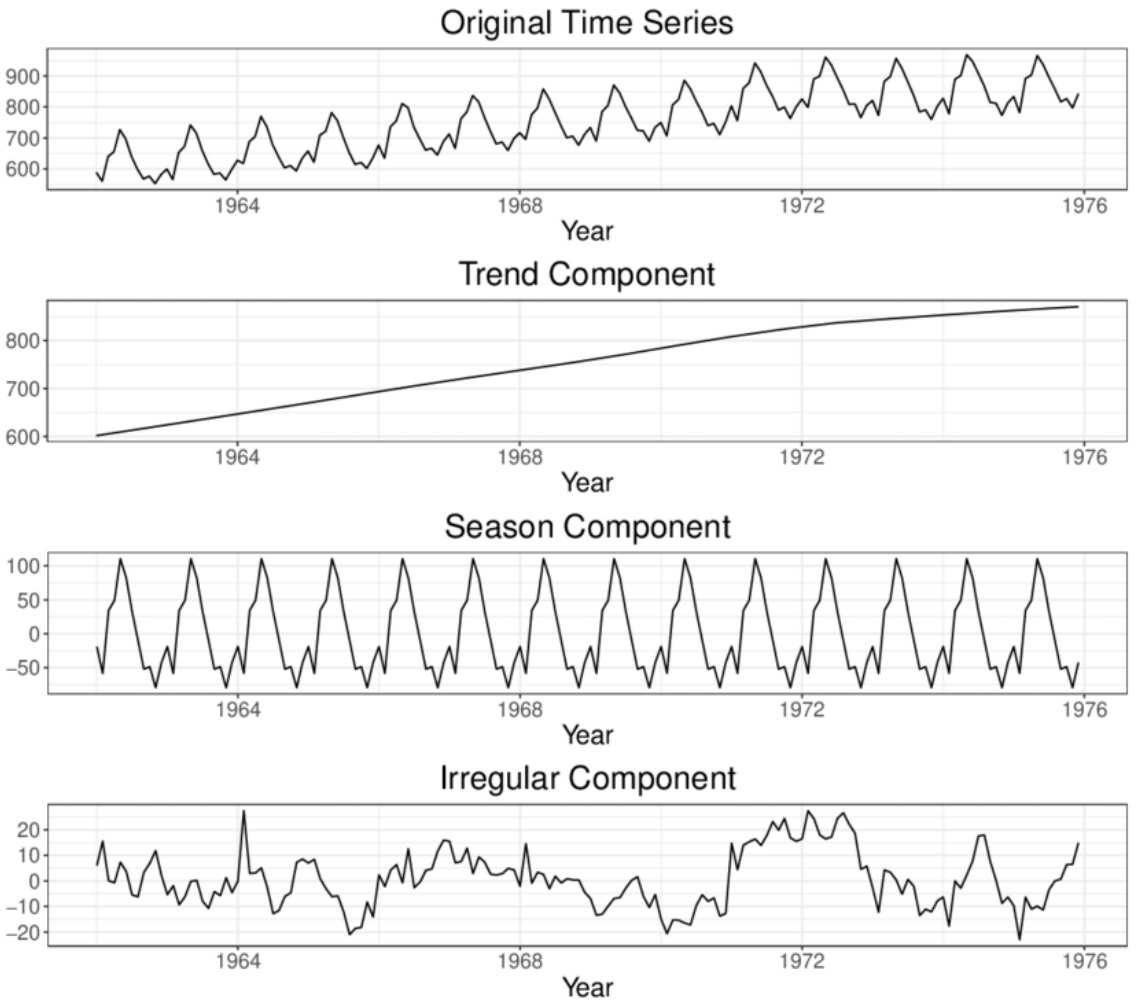
Autoformer: Decomposition Transformers with Auto-Correlation for Long-Term Series Forecasting

Мотивация:

- Увеличение горизонта предсказания временного ряда – важная задача для реальных применений.
- Помочь в «распутывании» длинных временные закономерностей может помочь использования разложения ряда на тренд и сезонную компоненту.
- Трансформеры хорошо себя показали на последовательностях и в частности временных, рядах, но они плохо масштабируются по длине в обычном случае и теряют в качестве если использовать sparse attention.

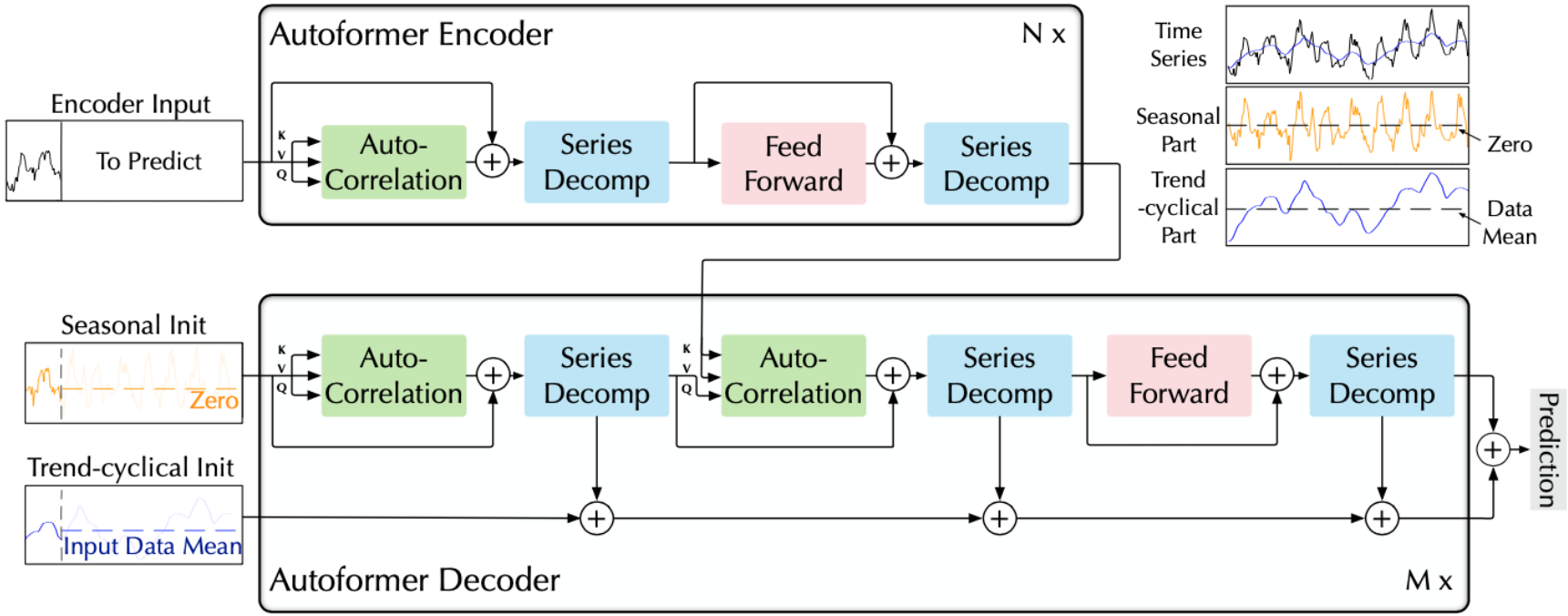
Time series decomposition

- Разложение исходного временного ряда на тренд, сезонность и остаток
- После разложения можно отдельно предсказывать тренд и сезонность с остатком, как более простые временные ряды



Пример разложения

Архитектура авторов – модифицированный трансформер



Series decomposition

$$\mathcal{X}_{\mathrm{t}} = \mathrm{AvgPool}(\mathrm{Padding}(\mathcal{X}))$$
$$\mathcal{X}_{\mathrm{s}} = \mathcal{X} - \mathcal{X}_{\mathrm{t}},$$

Autocorrelation layer

1. Коэффициент автокорреляции для периода τ

$$\mathcal{R}_{\mathcal{X}\mathcal{X}}(\tau) = \lim_{L\to\infty} \frac{1}{L}\sum_{t=1}^{L} \mathcal{X}_t \mathcal{X}_{t-\tau}.$$

2. Вычисление для всех периодов через преобразование Фурье

$$\mathcal{S}_{\mathcal{X}\mathcal{X}}(f) = \mathcal{F}(\mathcal{X}_t)\mathcal{F}^*(\mathcal{X}_t) = \int_{-\infty}^{\infty} \mathcal{X}_t e^{-i2\pi tf}\mathrm{d}t \overline{\int_{-\infty}^{\infty} \mathcal{X}_t e^{-i2\pi tf}\mathrm{d}t}$$

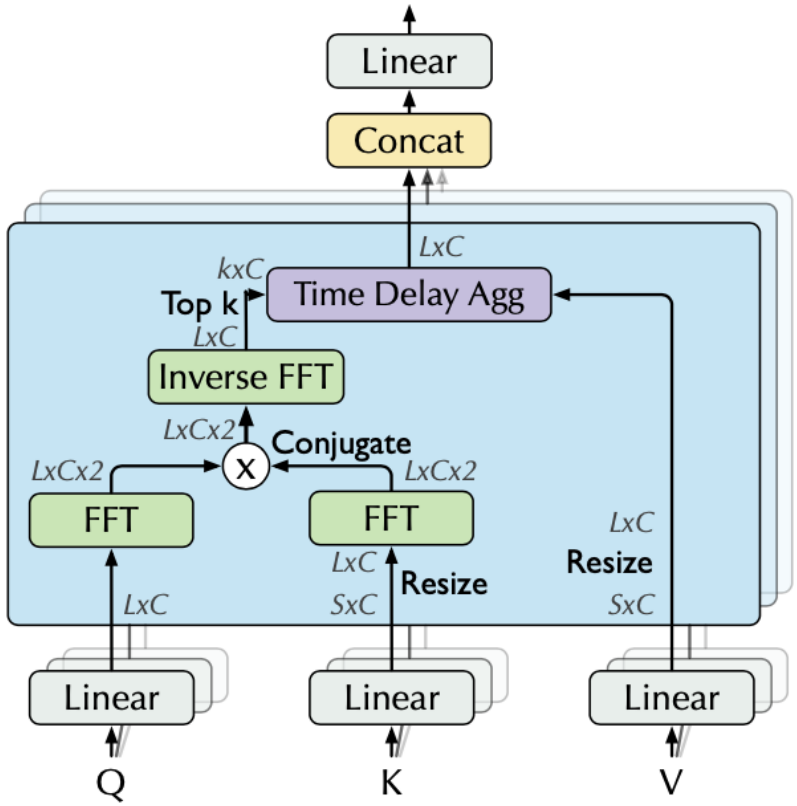
$$\mathcal{R}_{\mathcal{X}\mathcal{X}}(\tau) = \mathcal{F}^{-1}(\mathcal{S}_{\mathcal{X}\mathcal{X}}(f)) = \int_{-\infty}^{\infty} \mathcal{S}_{\mathcal{X}\mathcal{X}}(f) e^{i2\pi f\tau}\mathrm{d}f,$$

3. Time delay aggregation

$$\tau_1, \cdots, \tau_k = \underset{\tau\in\{1,\cdots,L\}}{\arg\mathrm{Topk}}\left(\mathcal{R}_{\mathcal{Q},\mathcal{K}}(\tau)\right)$$

$$\widehat{\mathcal{R}}_{\mathcal{Q},\mathcal{K}}(\tau_1), \cdots, \widehat{\mathcal{R}}_{\mathcal{Q},\mathcal{K}}(\tau_k) = \mathrm{SoftMax}\left(\mathcal{R}_{\mathcal{Q},\mathcal{K}}(\tau_1), \cdots, \mathcal{R}_{\mathcal{Q},\mathcal{K}}(\tau_k)\right)$$

$$\text{Auto-Correlation}(\mathcal{Q},\mathcal{K},\mathcal{V}) = \sum_{i=1}^{k} \mathrm{Roll}(\mathcal{V},\tau_i)\widehat{\mathcal{R}}_{\mathcal{Q},\mathcal{K}}(\tau_i),$$



Autocorrelation layer

Time delay aggregation

3. Time delay aggregation

$$\tau_1, \cdots, \tau_k = \underset{\tau \in \{1, \cdots, L\}}{\arg \operatorname{Topk}} \left(\mathcal{R}_{\mathcal{Q},\mathcal{K}}(\tau) \right)$$

$$\widehat{\mathcal{R}}_{\mathcal{Q},\mathcal{K}}(\tau_1), \cdots, \widehat{\mathcal{R}}_{\mathcal{Q},\mathcal{K}}(\tau_k) = \operatorname{SoftMax}\left(\mathcal{R}_{\mathcal{Q},\mathcal{K}}(\tau_1), \cdots, \mathcal{R}_{\mathcal{Q},\mathcal{K}}(\tau_k) \right)$$

$$\text{Auto-Correlation}(\mathcal{Q}, \mathcal{K}, \mathcal{V}) = \sum_{i=1}^{k} \operatorname{Roll}(\mathcal{V}, \tau_i) \widehat{\mathcal{R}}_{\mathcal{Q},\mathcal{K}}(\tau_i),$$

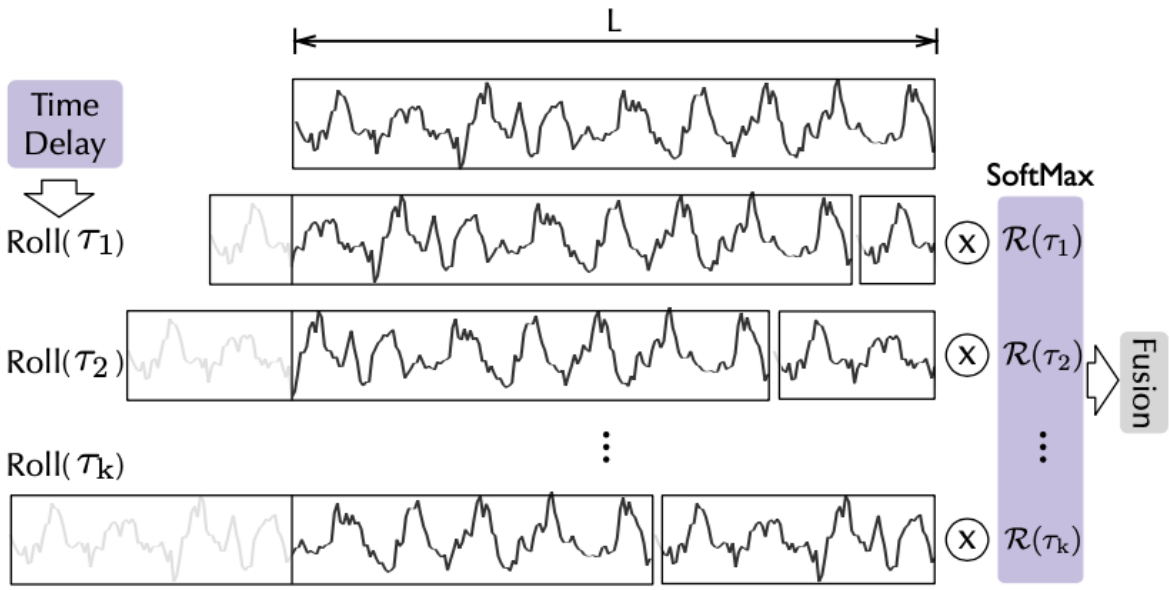
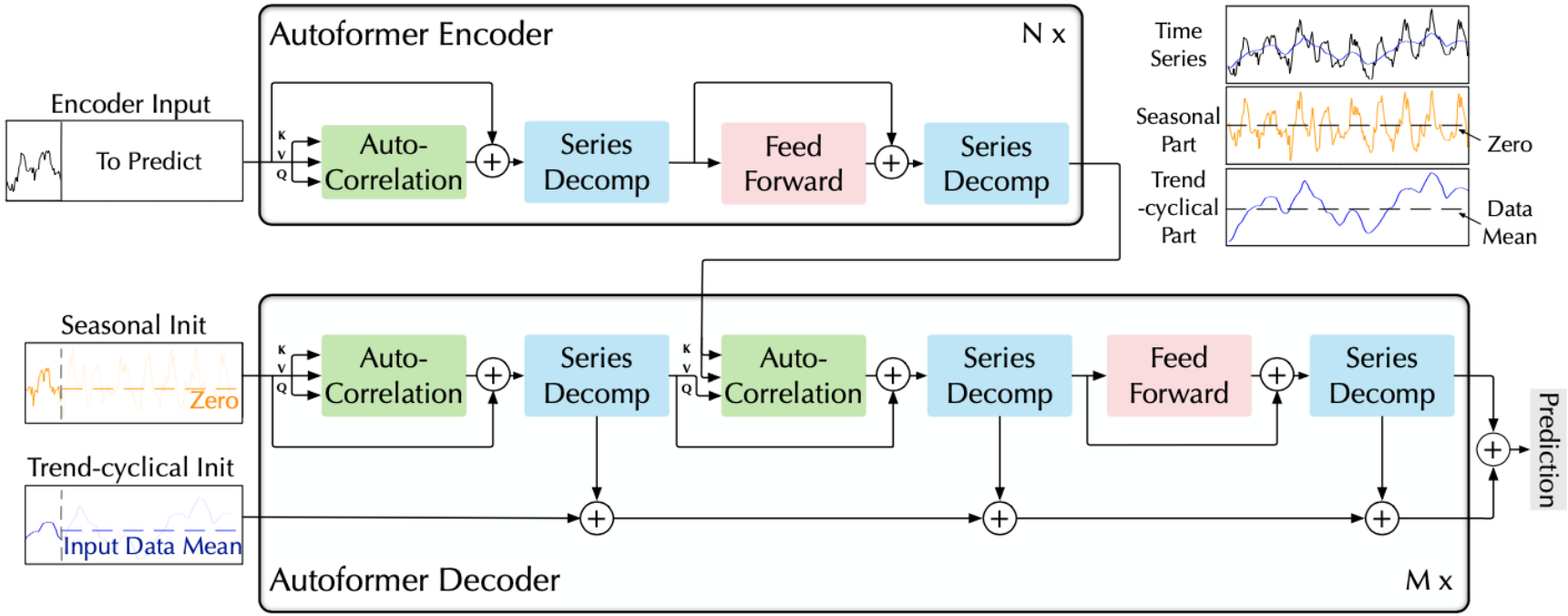


Схема time delay aggregation

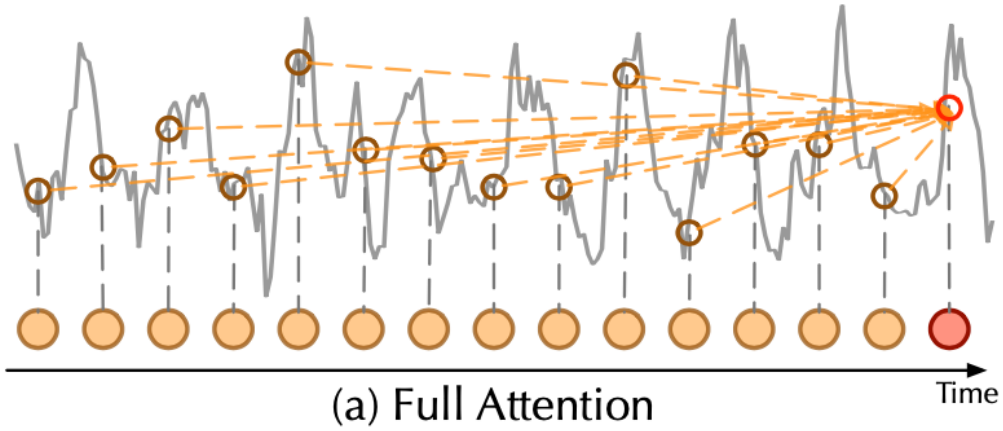
Архитектура авторов – модифицированный трансформер

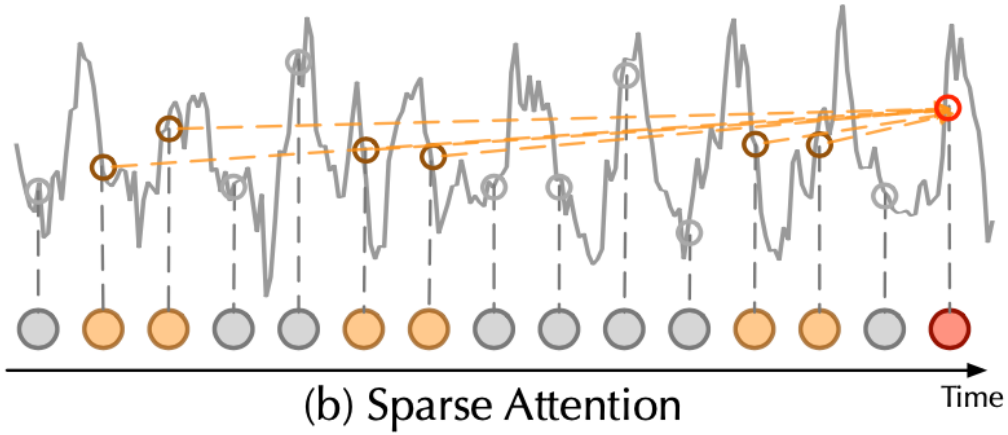


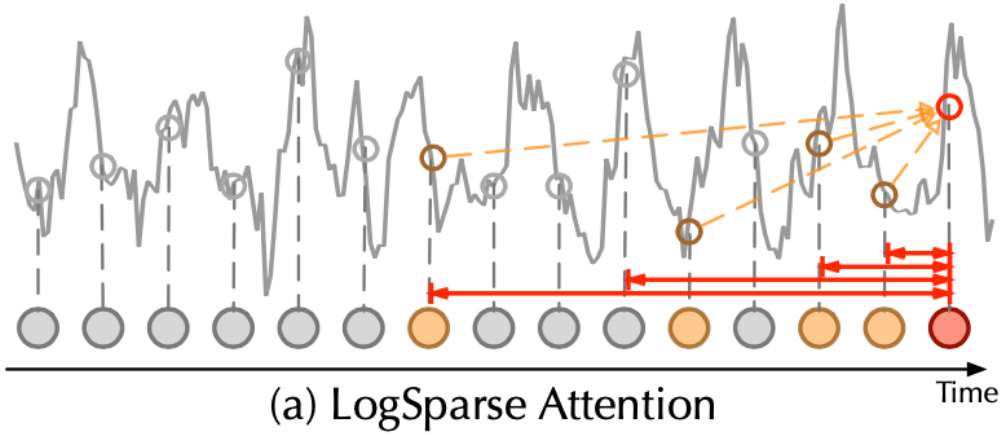
Series decomposition

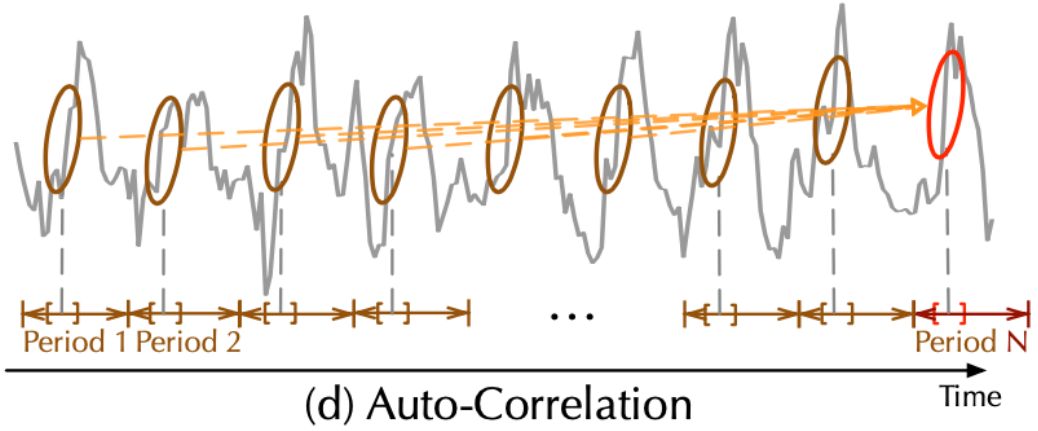
$$\mathcal{X}_{\mathrm{t}} = \mathrm{AvgPool}(\mathrm{Padding}(\mathcal{X}))$$
$$\mathcal{X}_{\mathrm{s}} = \mathcal{X} - \mathcal{X}_{\mathrm{t}},$$

Point-wise vs series-wise









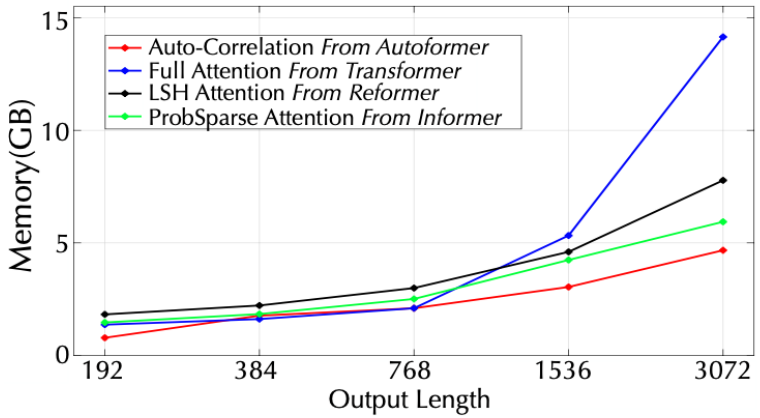
Результаты

1. EET: почасовое потребление электроэнергии с 2012 по 2014 год 321 потребителям
2. Electricity: температурные данные с трансформаторов каждые 15 минут
3. Exchange: обменные курсы восьми разных стран в период с 1990 по 2016 год.
4. Traffic: уровень загруженности различных автомобильных полос на автострадах в районе залива Сан-Франциско.
5. Weather: метеорологические данные за 2020 год (с разрешением в 10 минут) о температуре, влажности, etc
6. ILI: число заболевший простудными заболеваниями с 2002 по 2021 в одном из мед центров США (разрешение 1 неделя)

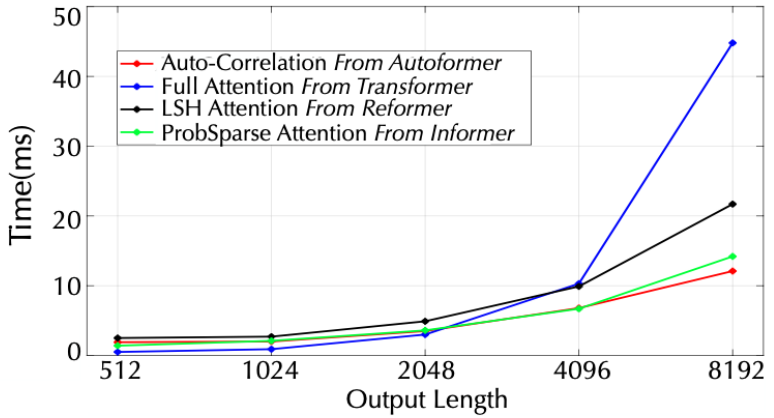
Models		**Autoformer**		Informer[41]		LogTrans[20]		Reformer[17]		LSTNet[19]		LSTM[13]		TCN[3]	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETT*	96	**0.255**	**0.339**	0.365	0.453	0.768	0.642	0.658	0.619	3.142	1.365	2.041	1.073	3.041	1.330
	192	**0.281**	**0.340**	0.533	0.563	0.989	0.757	1.078	0.827	3.154	1.369	2.249	1.112	3.072	1.339
	336	**0.339**	**0.372**	1.363	0.887	1.334	0.872	1.549	0.972	3.160	1.369	2.568	1.238	3.105	1.348
	720	**0.422**	**0.419**	3.379	1.388	3.048	1.328	2.631	1.242	3.171	1.368	2.720	1.287	3.135	1.354
Electricity	96	**0.201**	**0.317**	0.274	0.368	0.258	0.357	0.312	0.402	0.680	0.645	0.375	0.437	0.985	0.813
	192	**0.222**	**0.334**	0.296	0.386	0.266	0.368	0.348	0.433	0.725	0.676	0.442	0.473	0.996	0.821
	336	**0.231**	**0.338**	0.300	0.394	0.280	0.380	0.350	0.433	0.828	0.727	0.439	0.473	1.000	0.824
	720	**0.254**	**0.361**	0.373	0.439	0.283	0.376	0.340	0.420	0.957	0.811	0.980	0.814	1.438	0.784
Exchange	96	**0.197**	**0.323**	0.847	0.752	0.968	0.812	1.065	0.829	1.551	1.058	1.453	1.049	3.004	1.432
	192	**0.300**	**0.369**	1.204	0.895	1.040	0.851	1.188	0.906	1.477	1.028	1.846	1.179	3.048	1.444
	336	**0.509**	**0.524**	1.672	1.036	1.659	1.081	1.357	0.976	1.507	1.031	2.136	1.231	3.113	1.459
	720	**1.447**	**0.941**	2.478	1.310	1.941	1.127	1.510	1.016	2.285	1.243	2.984	1.427	3.150	1.458
Traffic	96	**0.613**	**0.388**	0.719	0.391	0.684	0.384	0.732	0.423	1.107	0.685	0.843	0.453	1.438	0.784
	192	**0.616**	**0.382**	0.696	0.379	0.685	0.390	0.733	0.420	1.157	0.706	0.847	0.453	1.463	0.794
	336	**0.622**	**0.337**	0.777	0.420	0.733	0.408	0.742	0.420	1.216	0.730	0.853	0.455	1.479	0.799
	720	**0.660**	**0.408**	0.864	0.472	0.717	0.396	0.755	0.423	1.481	0.805	1.500	0.805	1.499	0.804
Weather	96	**0.266**	**0.336**	0.300	0.384	0.458	0.490	0.689	0.596	0.594	0.587	0.369	0.406	0.615	0.589
	192	**0.307**	**0.367**	0.598	0.544	0.658	0.589	0.752	0.638	0.560	0.565	0.416	0.435	0.629	0.600
	336	**0.359**	**0.395**	0.578	0.523	0.797	0.652	0.639	0.596	0.597	0.587	0.455	0.454	0.639	0.608
	720	**0.419**	**0.428**	1.059	0.741	0.869	0.675	1.130	0.792	0.618	0.599	0.535	0.520	0.639	0.610
ILI	24	**3.483**	**1.287**	5.764	1.677	4.480	1.444	4.400	1.382	6.026	1.770	5.914	1.734	6.624	1.830
	36	**3.103**	**1.148**	4.755	1.467	4.799	1.467	4.783	1.448	5.340	1.668	6.631	1.845	6.858	1.879
	48	**2.669**	**1.085**	4.763	1.469	4.800	1.468	4.832	1.465	6.080	1.787	6.736	1.857	6.968	1.892
	60	**2.770**	**1.125**	5.264	1.564	5.278	1.560	4.882	1.483	5.548	1.720	6.870	1.879	7.127	1.918

Ablation studies. Сравнение с attention

Input Length I		96			192			336		
Prediction Length O		336	720	1440	336	720	1440	336	720	1440
Auto-Correlation	MSE	**0.339**	**0.422**	**0.555**	**0.355**	**0.429**	**0.503**	**0.361**	**0.425**	**0.574**
	MAE	**0.372**	**0.419**	**0.496**	**0.392**	**0.430**	**0.484**	**0.406**	**0.440**	**0.534**
Full Attention[35]	MSE	0.375	0.537	0.667	0.450	0.554	-	0.501	0.647	-
	MAE	0.425	0.502	0.589	0.470	0.533	-	0.485	0.491	-
LogSparse Attention[20]	MSE	0.362	0.539	0.582	0.420	0.552	0.958	0.474	0.601	-
	MAE	0.413	0.522	0.529	0.450	0.513	0.736	0.474	0.524	-
LSH Attention[17]	MSE	0.366	0.502	0.663	0.407	0.636	1.069	0.442	0.615	-
	MAE	0.404	0.475	0.567	0.421	0.571	0.756	0.476	0.532	-
ProbSparse Attention[41]	MSE	0.481	0.822	0.715	0.404	1.148	0.732	0.417	0.631	1.133
	MAE	0.472	0.559	0.586	0.425	0.654	0.602	0.434	0.528	0.691



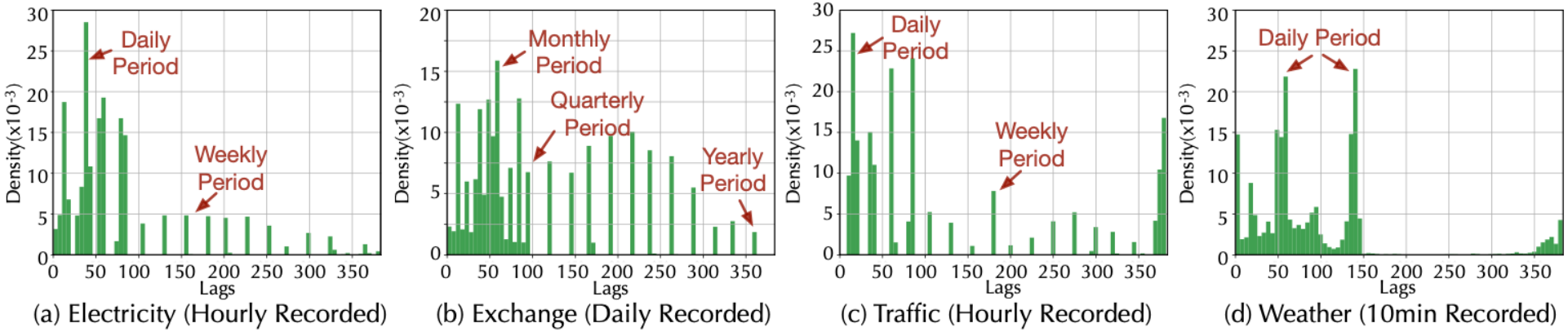
(a) Memory Efficiency Analysis



(b) Running Time Efficiency Analysis

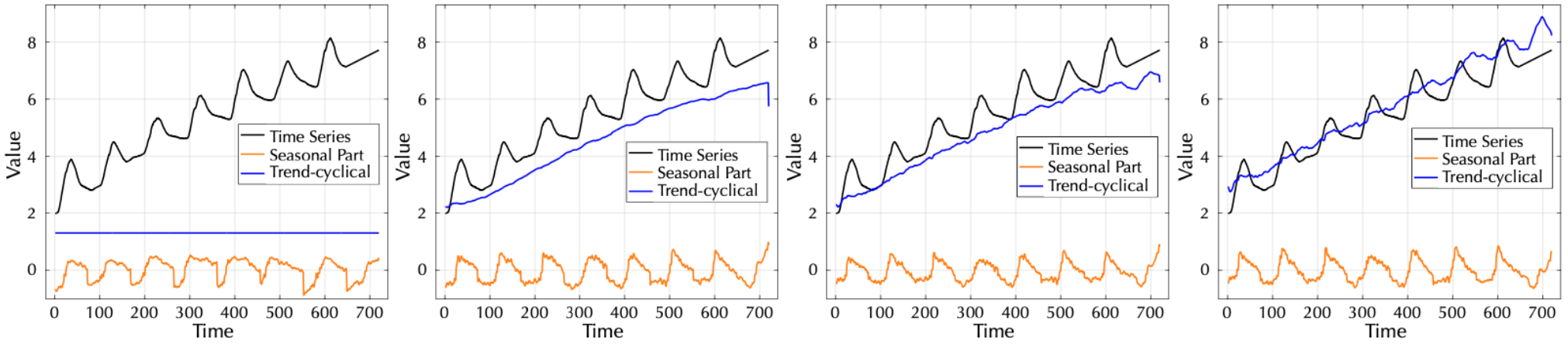
Model analysis

Распределение лагов auto-correlation на тесте



(a) Electricity (Hourly Recorded) (b) Exchange (Daily Recorded) (c) Traffic (Hourly Recorded) (d) Weather (10min Recorded)

Результаты в зависимости от числа блоков с series decomposition в декодере



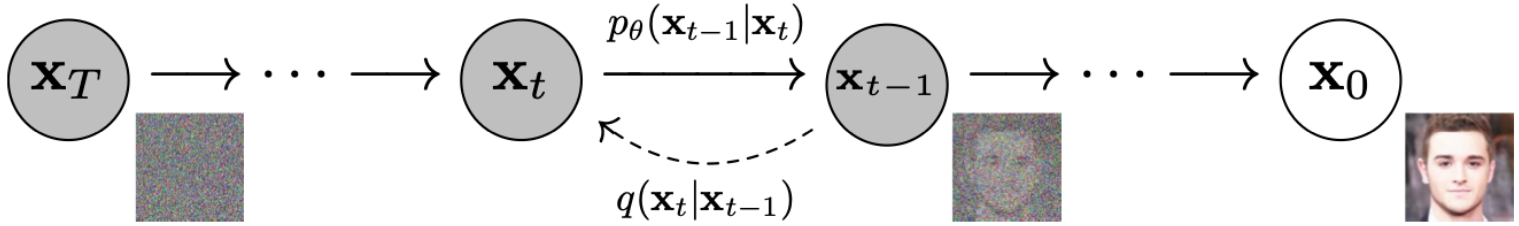
(a) Without decomposition block (b) One decomposition block (c) Two decomposition blocks (d) Three decomposition blocks

Autoregressive Denoising Diffusion Models for Multivariate Probabilistic Time Series Forecasting (ICML 2021)

Мотивация:

- Для учёта неопределённости предсказания хорошо бы уметь делать вероятностные прогнозы временных.
- Отдельные временные ряды могут быть взаимосвязаны и поэтому имеет смысл строить модели, которые принимают на вход много рядов и сразу делают предсказания для всех из них.

Теория диффузионных моделей кратко



Рассматриваемая графическая модель

$$p_\theta(\mathbf{x}_{0:T}) := p(\mathbf{x}_T) \prod_{t=1}^{T} p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t), \qquad p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) := \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t))$$

$$q(\mathbf{x}_{1:T}|\mathbf{x}_0) := \prod_{t=1}^{T} q(\mathbf{x}_t|\mathbf{x}_{t-1}), \qquad q(\mathbf{x}_t|\mathbf{x}_{t-1}) := \mathcal{N}(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t \mathbf{I})$$

$$\mathbb{E}\left[-\log p_\theta(\mathbf{x}_0)\right] \leq \mathbb{E}_q\left[-\log \frac{p_\theta(\mathbf{x}_{0:T})}{q(\mathbf{x}_{1:T}|\mathbf{x}_0)}\right] = \mathbb{E}_q\left[-\log p(\mathbf{x}_T) - \sum_{t\geq 1} \log \frac{p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)}{q(\mathbf{x}_t|\mathbf{x}_{t-1})}\right]$$

$$\mathbb{E}_q\Bigg[\underbrace{D_{\mathrm{KL}}(q(\mathbf{x}_T|\mathbf{x}_0) \parallel p(\mathbf{x}_T))}_{L_T} + \sum_{t>1} \underbrace{D_{\mathrm{KL}}(q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0) \parallel p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t))}_{L_{t-1}} \underbrace{-\log p_\theta(\mathbf{x}_0|\mathbf{x}_1)}_{L_0}\Bigg]$$

Теория диффузионных моделей кратко

$$\mathbb{E}_q \Bigg[\underbrace{D_{\mathrm{KL}}(q(\mathbf{x}_T|\mathbf{x}_0) \parallel p(\mathbf{x}_T))}_{L_T} + \sum_{t>1} \underbrace{D_{\mathrm{KL}}(q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0) \parallel p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t))}_{L_{t-1}} \underbrace{- \log p_\theta(\mathbf{x}_0|\mathbf{x}_1)}_{L_0} \Bigg]$$

ELBO

$$L_{t-1} = \mathbb{E}_q \left[\frac{1}{2\sigma_t^2} \|\tilde{\boldsymbol{\mu}}_t(\mathbf{x}_t, \mathbf{x}_0) - \boldsymbol{\mu}_\theta(\mathbf{x}_t, t)\|^2 \right] + C$$

$$\mathbb{E}_{\mathbf{x}_0, \boldsymbol{\epsilon}} \left[\frac{1}{2\sigma_t^2} \left\| \frac{1}{\sqrt{\alpha_t}} \left(\mathbf{x}_t(\mathbf{x}_0, \boldsymbol{\epsilon}) - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}} \boldsymbol{\epsilon} \right) - \boldsymbol{\mu}_\theta(\mathbf{x}_t(\mathbf{x}_0, \boldsymbol{\epsilon}), t) \right\|^2 \right]$$

$$\mathbb{E}_{\mathbf{x}_0, \boldsymbol{\epsilon}} \left[\frac{\beta_t^2}{2\sigma_t^2 \alpha_t (1-\bar{\alpha}_t)} \left\| \boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}, t) \right\|^2 \right]$$

Algorithm 1 Training

1: **repeat**
2: $\mathbf{x}_0 \sim q(\mathbf{x}_0)$
3: $t \sim \mathrm{Uniform}(\{1, \dots, T\})$
4: $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
5: Take gradient descent step on
$\nabla_\theta \left\| \boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}, t) \right\|^2$
6: **until** converged

Итоговый алгоритм обучения

Модель авторов = RNN + диффузионная модель

$$q_{\mathcal{X}}(\mathbf{x}^0_{t_0:T}|\mathbf{x}^0_{1:t_0-1}, \mathbf{c}_{1:T}) = \Pi_{t=t_0}^{T} q_{\mathcal{X}}(\mathbf{x}^0_t|\mathbf{x}^0_{1:t-1}, \mathbf{c}_{1:T})$$

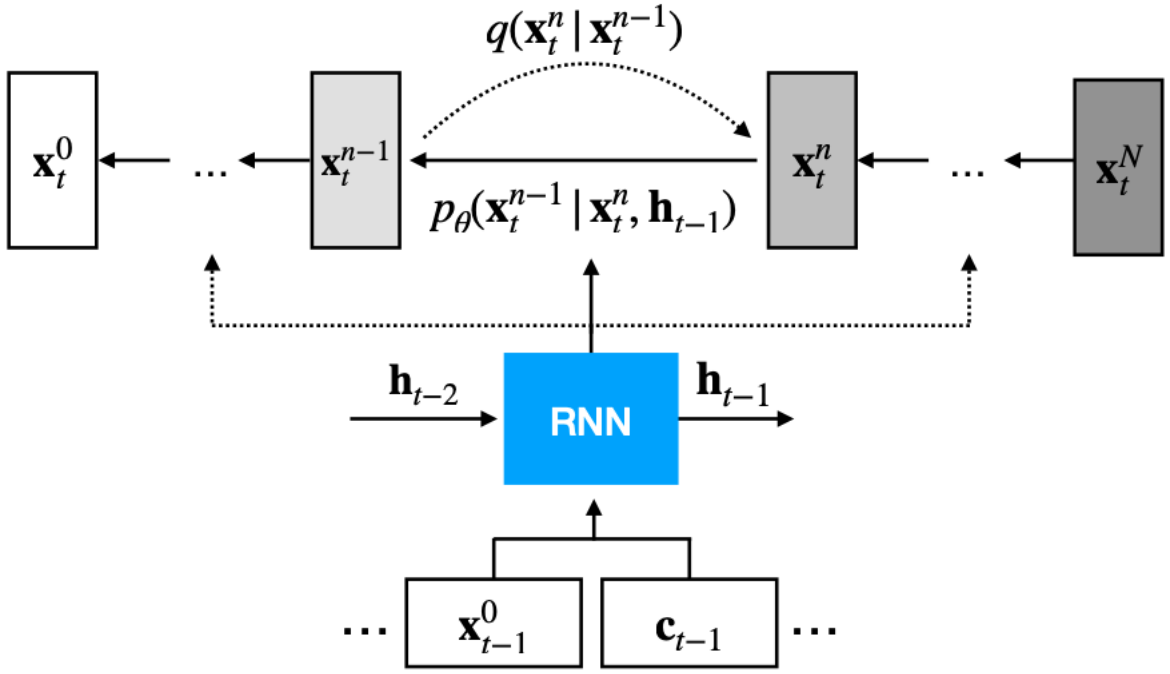


Схема модели

Algorithm 1 Training for each time series step $t \in [t_0, T]$

Input: data $\mathbf{x}_t^0 \sim q_{\mathcal{X}}(\mathbf{x}_t^0)$ and state $\mathbf{h}_{t-1}$
repeat
Initialize $n \sim \text{Uniform}(1, \ldots, N)$ and $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
Take gradient step on

$$\nabla_\theta \|\epsilon - \epsilon_\theta(\sqrt{\bar{\alpha}_n}\mathbf{x}_t^0 + \sqrt{1-\bar{\alpha}_n}\epsilon, \mathbf{h}_{t-1}, n)\|^2$$

until converged

Алгоритм обучения

Архитектура диффузионной части

- Состоит из 8 redisual блоков
- Номеру шага n в диффузионной модели сопоставляется эмбеддинг как в трансформере
- На входе x_t рассматривается как одномерная последовательность, в которой каждый элемент это элемент временного ряда в момент t.
- Dilated Conv1d в каждом блоке имеет поочерёдно dilation 1 и 2
- Gated act. unit – произведение сигмоиды от первой половины каналов на тангенс от второй половины каналов.

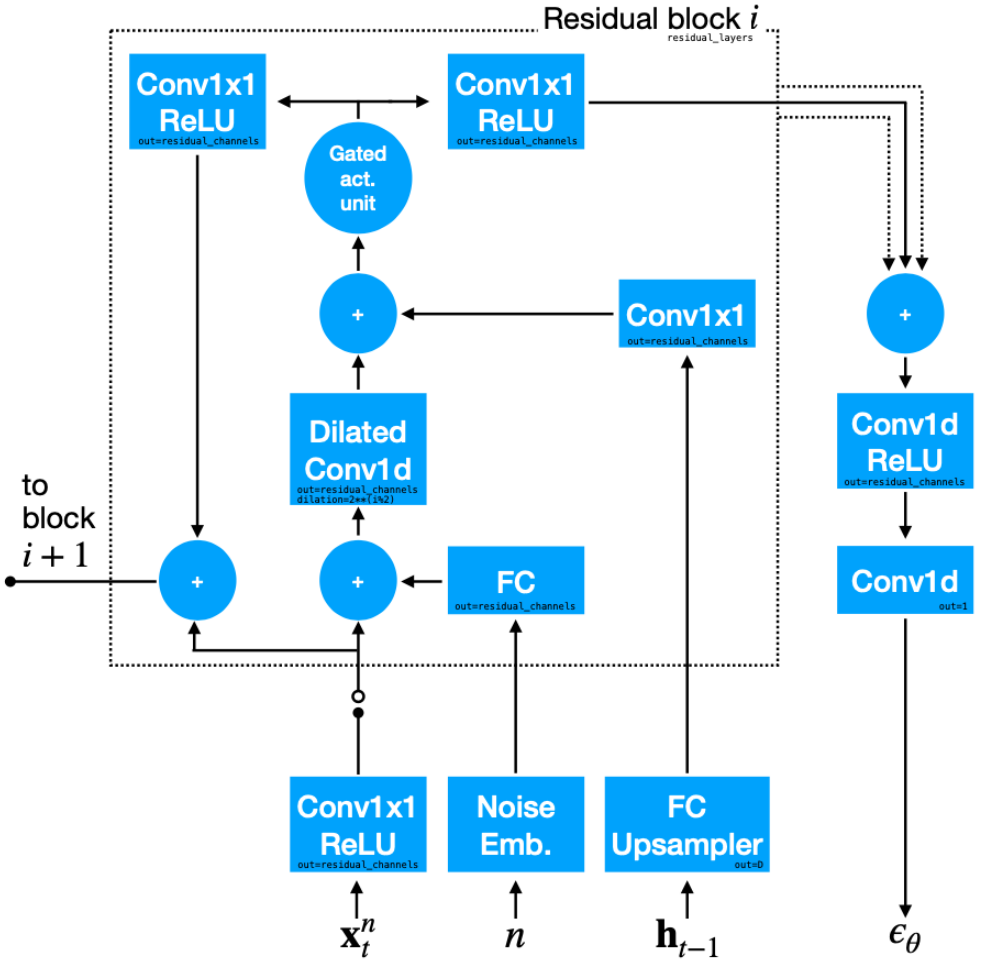


Схема residual блока

Датасеты и способ оценивания

Данные:

- Exchange - обменные курсы восьми разных стран в период с 1990 по 2016 год.
- Solar – выработка электрической энергии на солнечных панелях.
- Elec - почасовое потребление электроэнергии с 2012 по 2014 год 321 потребителям.
- Traffic - уровень загруженности различных автомобильных полос на автострадах в районе залива Сан-Франциско.

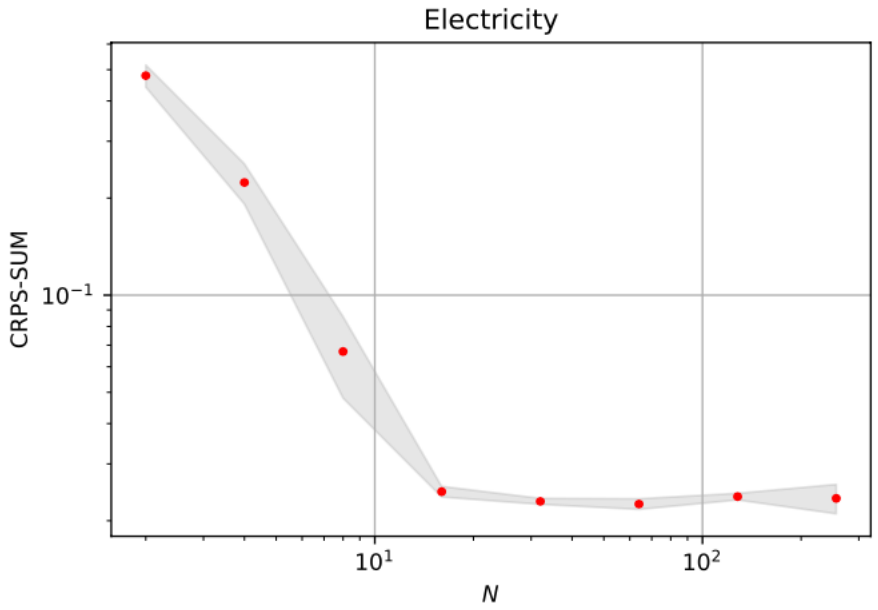
DATA SET	DIM. D	DOM.	FREQ.	TIME STEPS	PRED. STEPS
EXCHANGE	8	$\mathbb{R}^+$	DAY	6,071	30
SOLAR	137	$\mathbb{R}^+$	HOUR	7,009	24
ELEC.	370	$\mathbb{R}^+$	HOUR	5,833	24
TRAFFIC	963	$(0,1)$	HOUR	4,001	24
TAXI	1,214	$\mathbb{N}$	30-MIN	1,488	24
WIKI.	2,000	$\mathbb{N}$	DAY	792	30

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

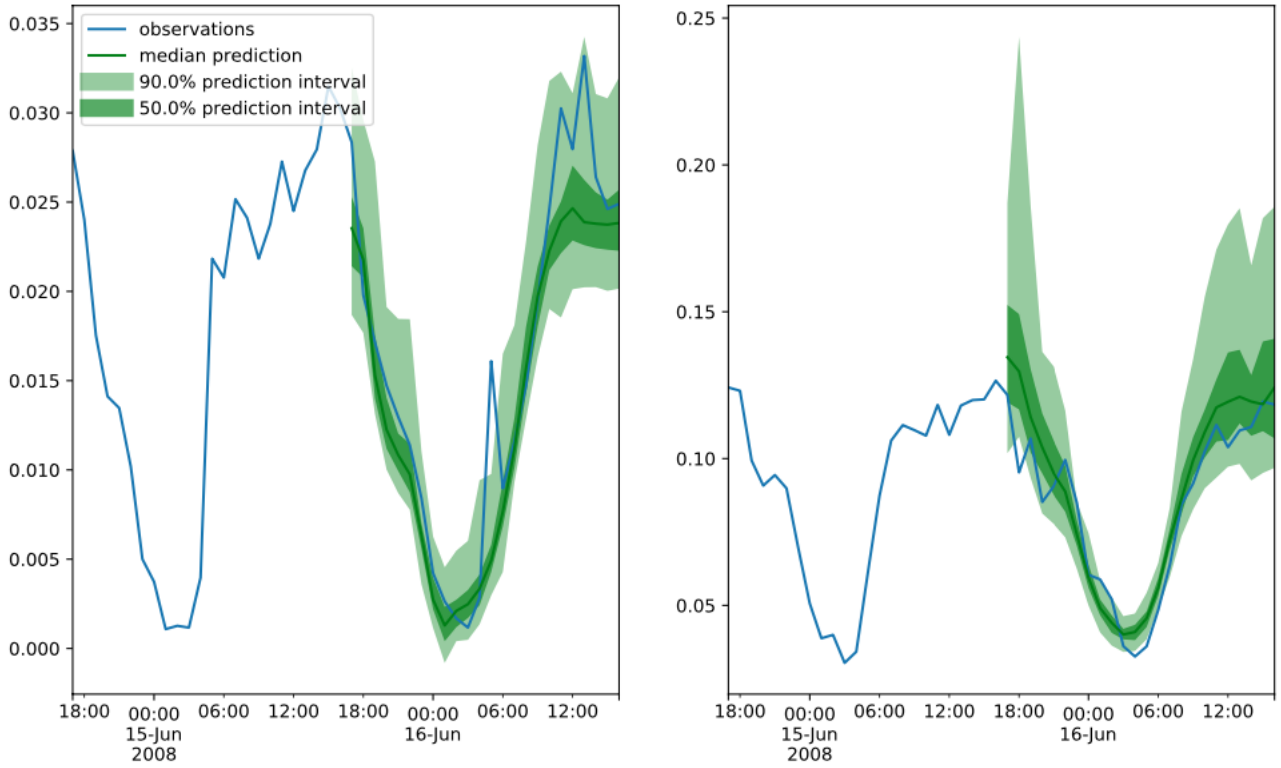
Результаты

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

Влияние числа шагов диффузионной модели и пример предсказания



Влияние числа шагов в диффузионной модели



Предсказания для пары временных рядов из датасета Traffic

Спасибо за внимание