NEURAL INFORMATION PROCESSING SYSTEMS | 2024

Task-oriented Time Series Imputation Evaluation via Generalized Representers

Zhixian Wang[1,2], Linxiao Yang[2], Sun Liang[2], Qingsong Wen[2], Yi Wang[1*]

[1] The University of Hong Kong [2] DAMO Academy, Alibaba Group

Zhixian Wang

Linxiao Yang

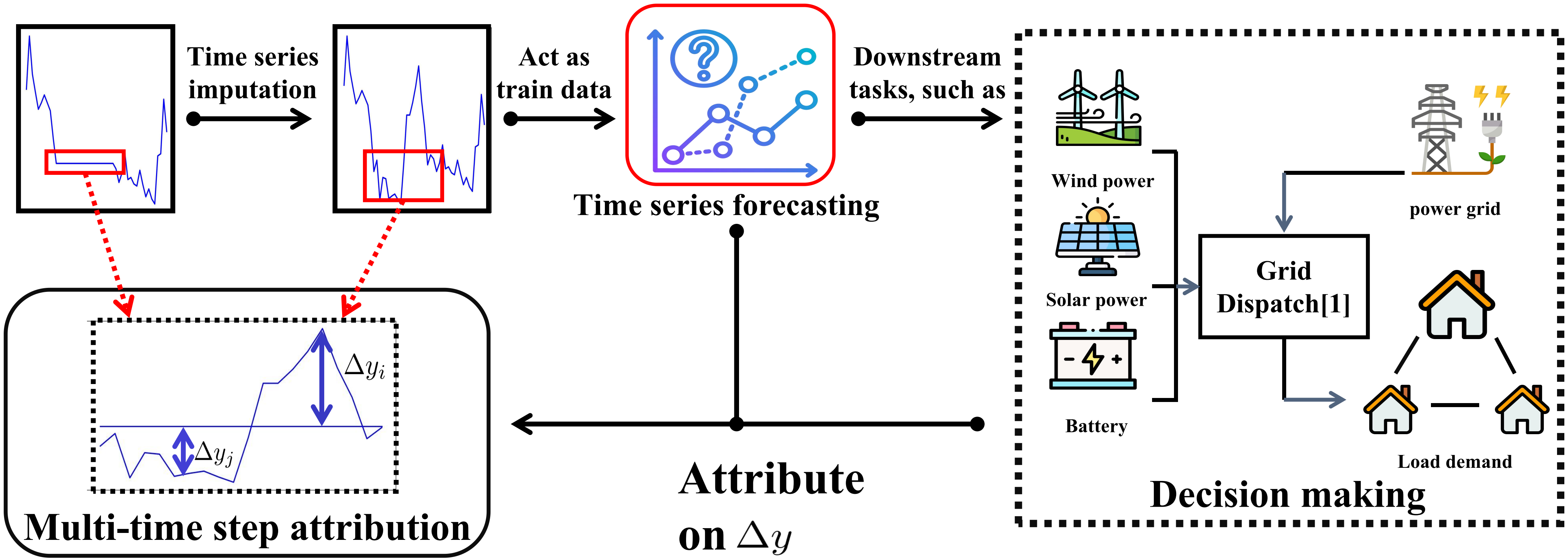
Liang Sun

Qingsong Wen

Yi Wang

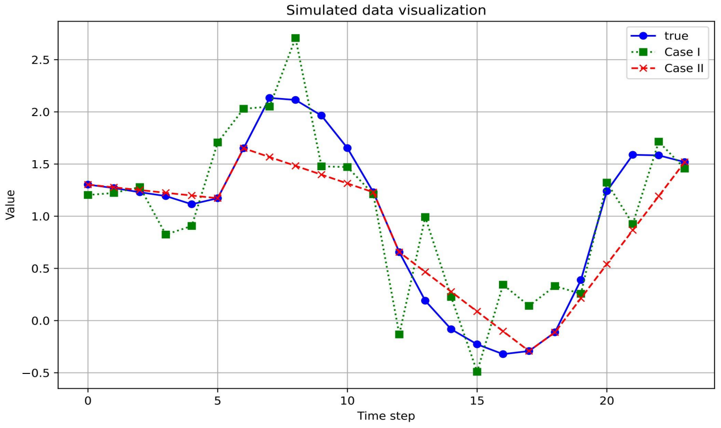
Introduction

- Time series imputation task can act as a prerequisite for other time series-related tasks.
- In time series forecasting task, time series data serves as both input data and training labels, which places high demands on time series imputation.



[1] Di Piazza A, Di Piazza M C, La Tona G, et al. An artificial neural network-based forecasting model of energy-related time series for electrical grid management[J]. Mathematics and Computers in Simulation, 2021, 184: 294-305.

Introduction



MSE↓	Imputation	Forecasting
I	0.1039	**0.1140**
II	**0.0576**	0.1395

Toy example

Method	Air		PhysioNet2012		ETTm1	
	MAE ↓	MSE ↓	MAE ↓	MSE ↓	MAE ↓	MSE ↓
Mean	0.692±0.000	0.970±0.000	0.702±0.000	0.954±0.000	0.663±0.000	0.809±0.000
Median	0.660±0.000	1.027±0.000	0.685±0.000	0.991±0.000	0.657±0.000	0.825±0.000
LOCF	0.206±0.000	0.279±0.000	0.411±0.000	0.569±0.000	0.135±0.000	0.072±0.000
M-RNN	0.524±0.001	0.648±0.003	0.674±0.001	0.864±0.002	0.651±0.060	1.074±0.120
GP-VAE	0.280±0.003	0.266±0.009	0.400±0.007	0.433±0.011	0.290±0.017	0.178±0.015
BRITS	0.142±0.001	0.129±0.001	0.246±0.001	0.325±0.002	0.124±0.002	0.046±0.002
USGAN	0.141±0.001	0.132±0.001	0.250±0.001	0.306±0.001	0.127±0.005	0.048±0.003
CSDI	**0.105±0.003**	0.153±0.021	0.211±0.003	0.260±0.050	0.157±0.052	0.292±0.456
TimesNet	0.159±0.002	0.172±0.003	0.266±0.007	0.272±0.006	**0.113±0.006**	**0.027±0.002**
Transformer	0.163±0.003	0.160±0.004	0.209±0.002	0.225±0.002	0.133±0.009	0.035±0.004
SAITS	0.133±0.002	**0.128±0.001**	**0.202±0.002**	**0.218±0.002**	0.115±0.011	0.030±0.006

Table 3 in [1]

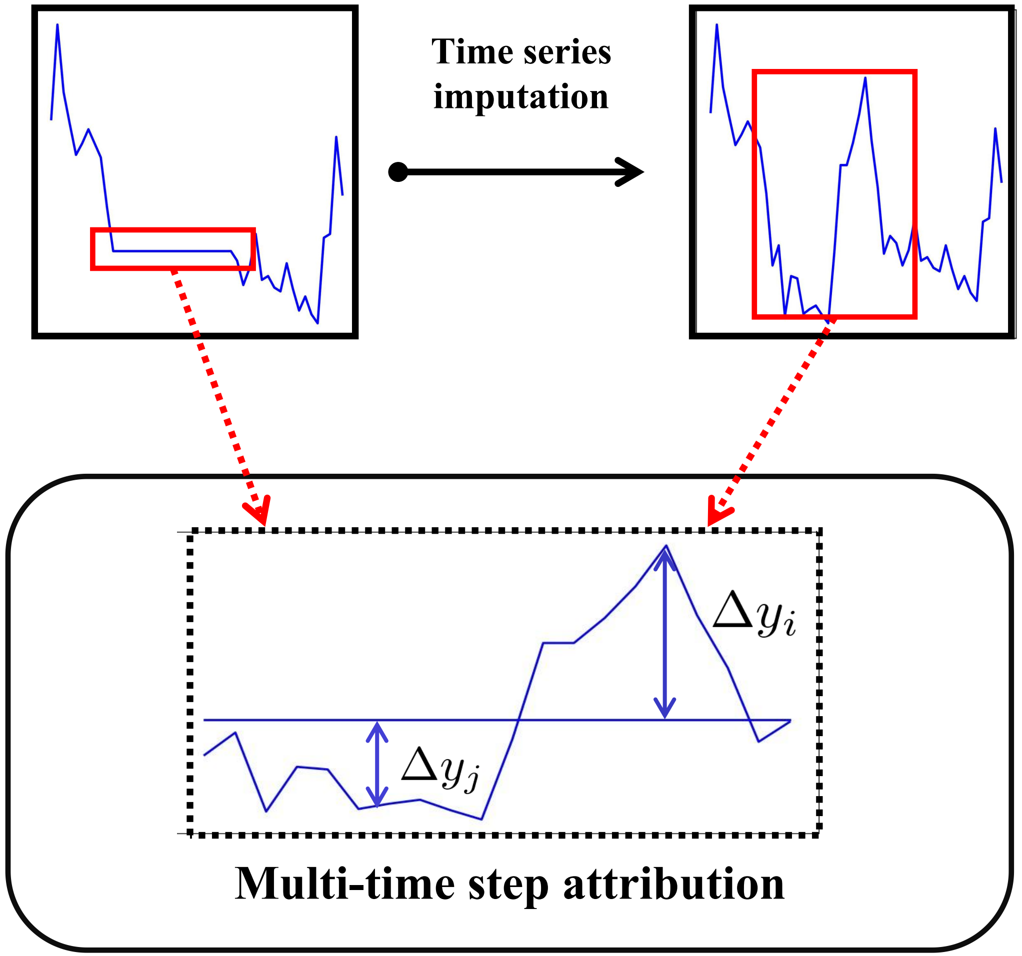
- ☐ The accuracy of time series imputation may not necessarily reflect the accuracy of its application in downstream forecasting tasks.

- ☐ There is currently no universal time imputation method that can outperform other methods on all datasets and evaluation metrics.

[1] Wang J, Du W, Cao W, et al. Deep Learning for Multivariate Time Series Imputation: A Survey[J].

Method

Problem Statement



$$I(i,l) = \sum_{k=1}^{m} I\left(i,l,\boldsymbol{X}_k^v\right) = \sum_{k=1}^{m} \left(\mathcal{L}\left(f\left(\boldsymbol{X}_k^v,\boldsymbol{\theta}_1\right),\boldsymbol{y}_k^v\right) - \mathcal{L}\left(f\left(\boldsymbol{X}_k^v,\boldsymbol{\theta}_2\right),\boldsymbol{y}_k^v\right)\right)$$

$$\text{s.t.}\ \boldsymbol{\theta}_1 = \arg\min_{\boldsymbol{\theta}} \sum_{k=1}^{n} \mathcal{L}\left(f\left(\boldsymbol{X}_k,\boldsymbol{\theta}\right),\boldsymbol{y}_k^{(1)}\right)$$

$$\boldsymbol{\theta}_2 = \arg\min_{\boldsymbol{\theta}} \mathcal{L}\left(f\left(\boldsymbol{X}_i,\boldsymbol{\theta}\right),\overline{\boldsymbol{y}}_i^{(2,l)}\right) + \sum_{k\neq i}^{n} \mathcal{L}\left(f\left(\boldsymbol{X}_k,\boldsymbol{\theta}\right),\boldsymbol{y}_k^{(1)}\right).$$

- **Task-oriented evaluation**
- **Without retraining**

Method

☐ First-order approximation

$$I(i,l) \approx \sum_{k=1}^{m} \left. \frac{\partial \mathcal{L}\left(f\left(\boldsymbol{X}_k^v, \boldsymbol{\theta}\right), \boldsymbol{y}_k^v\right)}{\partial y_{i,l}} \right|_{y_{i,l}=y_{i,l}^{(1)}} \left(y_{i,l}^{(1)} - y_{i,l}^{(2)}\right)$$

$$= \sum_{k=1}^{m} \frac{\partial \mathcal{L}\left(f\left(\boldsymbol{X}_k^v, \boldsymbol{\theta}\right), \boldsymbol{y}_k^v\right)^T}{\partial f\left(\boldsymbol{X}_k^v, \boldsymbol{\theta}\right)} \left. \frac{\partial f\left(\boldsymbol{X}_k^v, \boldsymbol{\theta}\right)}{\partial y_{i,l}} \right|_{y_{i,l}=y_{i,l}^{(1)}} \left(y_{i,l}^{(1)} - y_{i,l}^{(2)}\right).$$

☐ Kernel-machine approximation

$$\hat{\boldsymbol{\alpha}} = \underset{\boldsymbol{\alpha} \in \mathbb{R}^n \times \mathbb{R}^{L_2} \times \mathbb{R}^{L_2}}{\operatorname{argmin}} \left\{ \sum_{i=1}^{n} \sum_{l=1}^{L_2} \sum_{j=1}^{n} \mathcal{L}\left(\boldsymbol{\alpha}_{i,l}^T K\left(\boldsymbol{X}_i, \boldsymbol{X}_j\right), \frac{\partial f\left(\boldsymbol{X}_j, \boldsymbol{\theta}\right)}{\partial y_{i,l}} \right) \right\}.$$

Method

Intuition

Remark 1. *Given two infinitely differentiable functions $f(\boldsymbol{x})$ and $g(\boldsymbol{x})$ in a bounded domain $D \in R^n$, $||f(\boldsymbol{x}) - g(\boldsymbol{x})||$ is always less than ϵ. For any given δ and ϵ_2, there exists an ϵ such that, in the domain D, the measure of the region I that satisfying $||\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}} - \frac{\partial g(\boldsymbol{x})}{\partial \boldsymbol{x}}|| > \delta$ is not greater than ϵ_2, i.e, $m(I) \leq \epsilon_2$.*

Discussion

Use $g\left(X_{\text{test}}, y_{i,l}\right)$ **to approximate** $f\left(X_{\text{test}}, \theta_T, y_{i,l}\right)$ ⇨ **Use** $\frac{\partial g\left(X_{\text{test}}, y_{i,l}\right)}{\partial y_{i,l}}$ **to approximate** $\frac{\partial f\left(X_{\text{test}}, \theta_T, y_{i,l}\right)}{\partial y_{i,l}}$

⇨

$$\frac{\partial f\left(X_{\text{test}}, \theta_T, y_{i,l}\right)}{\partial y_{i,l}} = \frac{\partial f\left(X_{\text{test}}, \theta_T, y_{i,l}\right)}{\partial \theta_T} \frac{\partial \theta_T}{\partial y_{i,l}}$$

Constant

Method

Discussion

$$\frac{\partial f\left(X_{\text{test}}, \theta_T, y_{i,l}\right)}{\partial y_{i,l}} = \frac{\partial f\left(X_{\text{test}}, \theta_T, y_{i,l}\right)}{\partial \theta_T} \boxed{\frac{\partial \theta_T}{\partial y_{i,l}}}$$

(Let $\frac{\partial \mathcal{L}\left(f\left(X, \theta_t, y_{i,l}\right), y\right)}{\partial \theta_t}$ **be** $h_t\left(y_{i,l}\right)$**, and** $\frac{\partial^2 h_t\left(y_{i,l}\right)}{\partial y_{i,l}^2} = 0$**)**

For SGD

$$\theta_T = \theta_0 - \sum_{t=1}^{T} \eta h_t\left(y_{i,l}\right) \Rightarrow \frac{\partial^2 \theta_T}{\partial y_{i,l}^2} = 0$$

For Adam

$\frac{\partial \theta_T}{\partial y_{i,l}}$ **will be an algebraic function with only a finite number of monotonic intervals.**

Final Approximation Problem

$$\hat{\boldsymbol{\alpha}}' = \underset{\boldsymbol{\alpha}' \in \mathbb{R}^n \times \mathbb{R}^{L_2}}{\operatorname{argmin}} \left\{ \sum_{i=1}^{n} \mathcal{L}\left(\sum_{j=1}^{n} \boldsymbol{\alpha}_j'^T K\left(\boldsymbol{X}_i, \boldsymbol{X}_j\right), f\left(\boldsymbol{X}_i, \boldsymbol{\theta}\right) \right) \right\},$$

$$\hat{\boldsymbol{\alpha}_{i,l}} = \frac{\partial \hat{\boldsymbol{\alpha}}_i'}{\partial y_{i,l}}$$

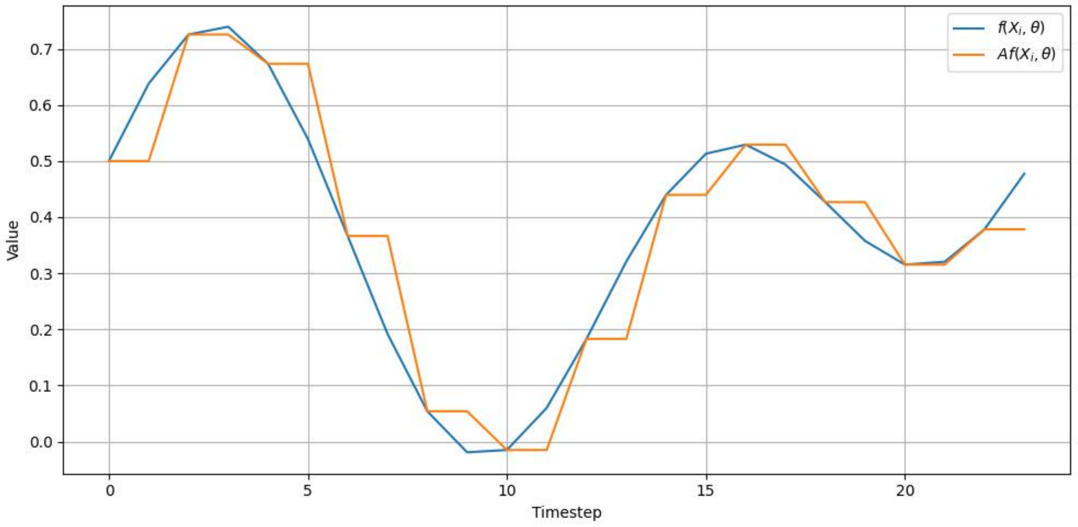
Method

Approximation Result

$$\sum_{k=1}^{m} -\frac{1}{n} \frac{\partial \mathcal{L}\left(f\left(\boldsymbol{X}_k^v, \boldsymbol{\theta}\right), \boldsymbol{y}_k^v\right)}{\partial f\left(\boldsymbol{X}_k^v, \boldsymbol{\theta}\right)} \underbrace{\frac{\partial^2 \mathcal{L}\left(f\left(\boldsymbol{X}_i, \boldsymbol{\theta}\right), \boldsymbol{y}_i\right)^T}{\partial f\left(\boldsymbol{X}_i, \boldsymbol{\theta}_t\right) \partial y_{i,l}}}_{\boldsymbol{a}_{\hat{i},l}} \underbrace{\frac{\partial f\left(\boldsymbol{X}_i, \boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}} \frac{\partial f\left(\boldsymbol{X}_k^v, \boldsymbol{\theta}\right)^T}{\partial \boldsymbol{\theta}}}_{\text{NTKkernel}}.$$

Accleration method

$$\frac{\partial f\left(\boldsymbol{X}_i, \boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}} \approx \boldsymbol{A}^{\dagger} \frac{\partial \boldsymbol{A} f\left(\boldsymbol{X}_i, \boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}}.$$



Experiment

☐ Imputation Evaluation

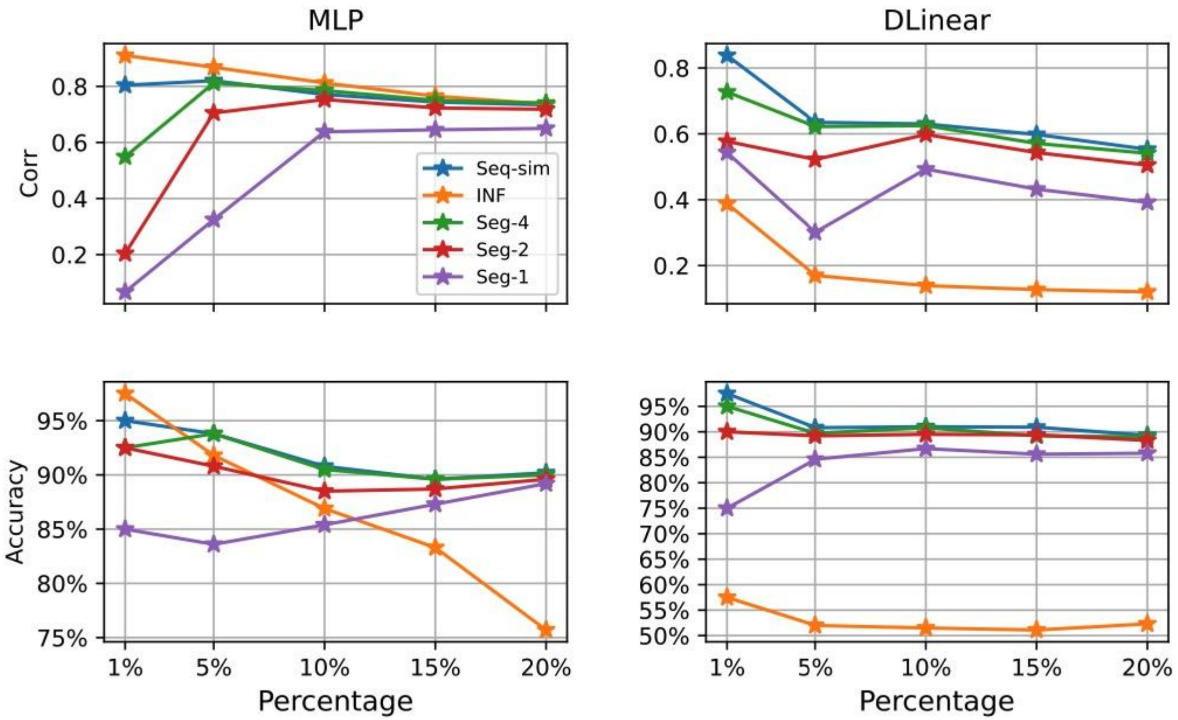


Fig. Comparision with true gain

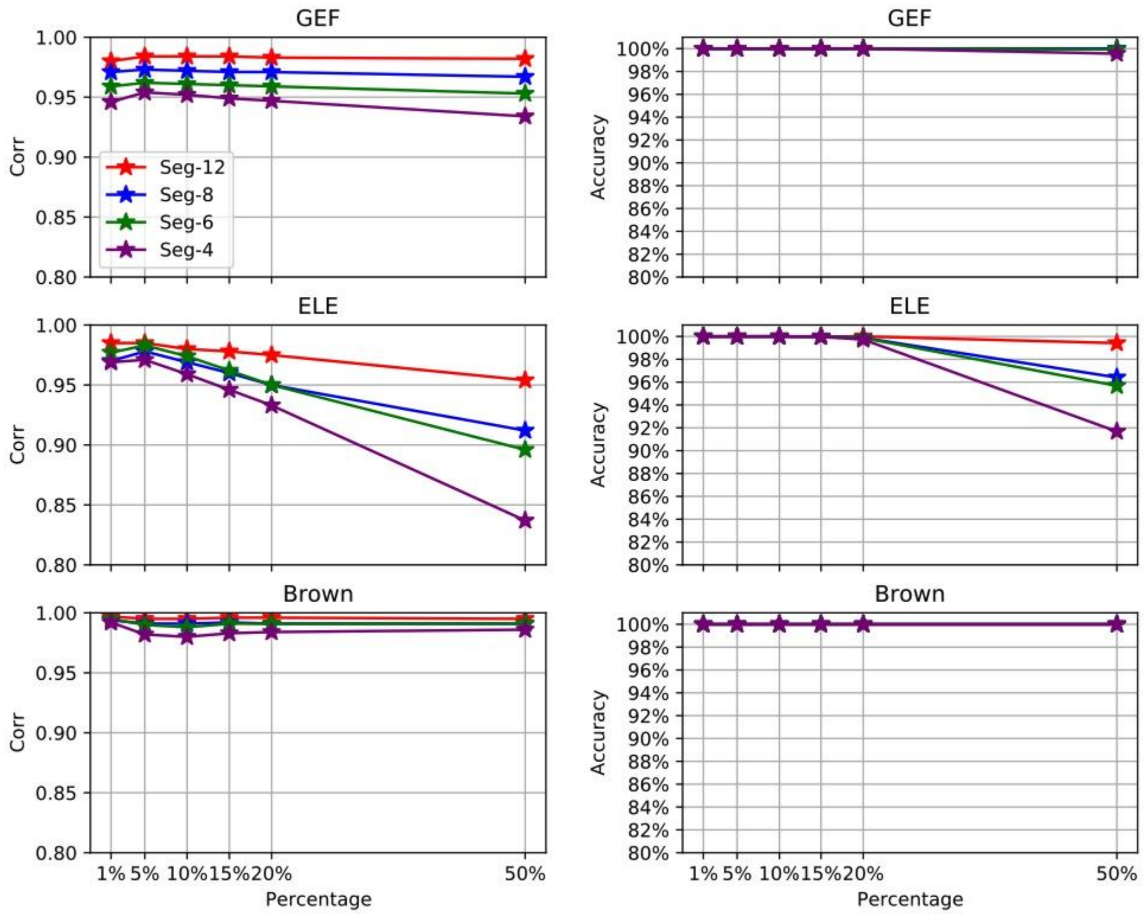


Fig. Comparision with accleration method

- ➢Corr: The correlation between estimated gains and retraining gains.
- ➢Accuracy: Accuracy of sign estimation for retraining gains.

Experiment

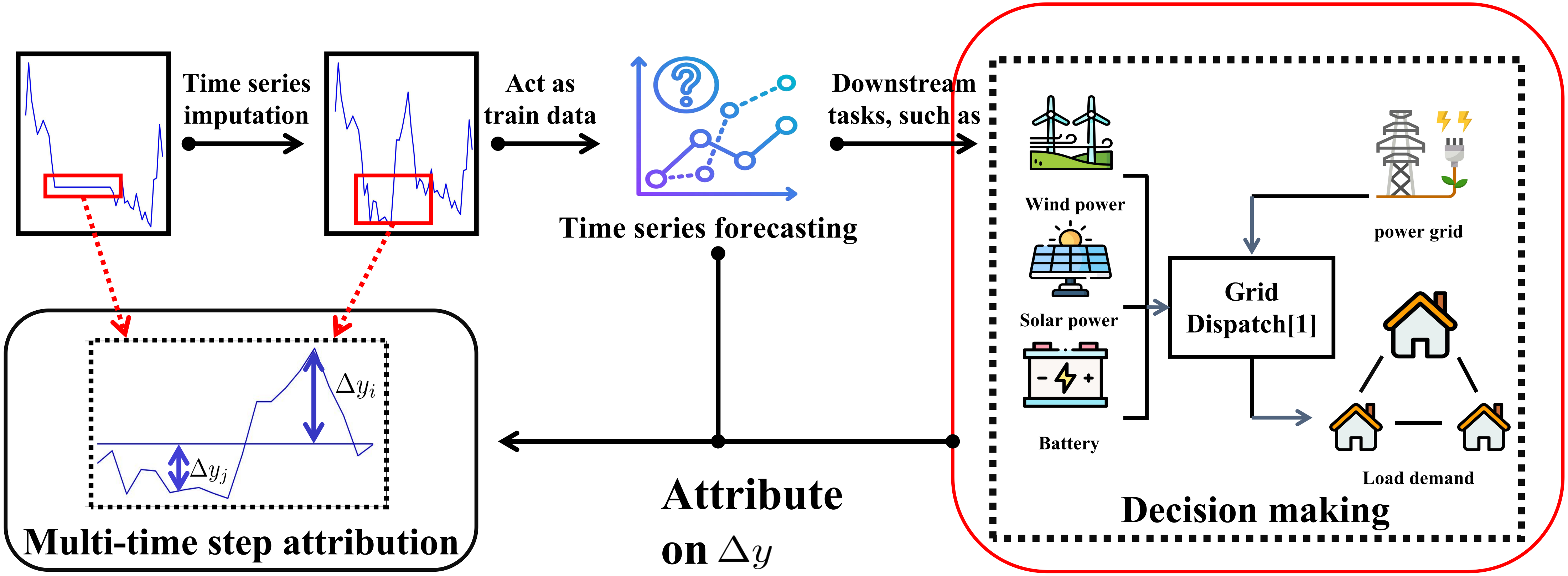
☐ MSE↓in the downstream forecasting task

Method	GEF	ETTH1	ETTH2	ELECTRICITY	TRAFFIC	AIR
I.Original						
Mean	0.1750	0.0523	0.1797	0.1123	0.4359	0.1508
SAITS	0.1980(0.0092)	0.1027(0.0021)	0.2098(0.0125)	0.1176(0.0110)	0.4311(0.0151)	0.5006(0.0251)
BRITS	0.2021(0.0007)	0.1692(0.0105)	0.2384(0.0018)	0.1503(0.0003)	0.4535(0.0001)	0.6979(0.0086)
MRNN	0.2052(0.0001)	0.2184(0.0016)	0.2317(0.0001)	-	0.4540(0.0000)	0.7965(0.0018)
GPVAE	0.2087(0.0019)	0.1591(0.0072)	0.2365(0.0022)	0.1471(0.0001)	0.4465(0.0001)	0.6968(0.0044)
USGAN	0.2048(0.0023)	0.1549(0.0179)	0.2238(0.0085)	0.1447(0.0011)	0.4742(0.0048)	0.6840(0.0306)
SPIN	0.2120(0.0029)	0.2000(0.0509)	0.2414(0.0327)	0.1588(0.0113)	0.4609(0.0148)	0.6604(0.0802)
ImputeFormer	0.1820(0.0016)	0.1558(0.0033)	0.2125(0.0022)	0.1076(0.0012)	0.4249(0.0060)	0.6300(0.0119)
II.With Gain estimation						
Mean+SAITS	**0.1653(0.0008)**	**0.0522(0.0000)**	0.1797(0.0000)	**0.0957(0.0006)**	**0.4147(0.0023)**	**0.1491(0.0001)**
Mean+BRITS	0.1694(0.0000)	**0.0522(0.0000)**	0.1795(0.0000)	0.1068(0.0000)	0.4318(0.0000)	0.1507(0.0000)
Mean+MRNN	0.1696(0.0000)	0.0523(0.0000)	0.1794(0.0000)	-	0.4319(0.0000)	0.1508(0.0000)
Mean+GPVAE	0.1696(0.0000)	**0.0522(0.0000)**	0.1795(0.0000)	0.1058(0.0001)	0.4290(0.0005)	0.1507(0.0000)
Mean+USGAN	0.1698(0.0001)	**0.0522(0.0000)**	0.1795(0.0000)	0.1069(0.0003)	0.4215(0.0004)	0.1506(0.0000)
Mean+SPIN	0.1679(0.0016)	0.0523(0.0001)	**0.1784(0.0000)**	0.1038(0.0007)	0.4276(0.0013)	0.1502(0.0005)
Mean+ImputeFormer	0.1657(0.0003)	**0.0522(0.0000)**	0.1795(0.0000)	0.0977(0.0002)	0.4178(0.0015)	0.1498(0.0000)
III.With Influence Function						
SATIS+INF	0.1953(0.0008)	0.1026(0.0021)	0.2074(0.0115)	0.1170(0.0169)	0.4294(0.0153)	0.5207(0.0213)
BRITS+INF	0.1952(0.0009)	0.1637(0.0091)	0.2326(0.0005)	0.1302(0.0022)	0.4419(0.0008)	0.7110(0.0069)
MRNN+INF	0.1972(0.0002)	0.1905(0.0017)	0.2251(0.0003)	-	0.4431(0.0002)	0.7758(0.0020)
GPVAE+INF	0.2013(0.0018)	0.1543(0.0073)	0.2314(0.0031)	0.1275(0.0027)	0.4347(0.0005)	0.7096(0.0021)
USGAN+INF	0.1984(0.0024)	0.1486(0.0120)	0.2191(0.0060)	0.1263(0.0013)	0.4597(0.0045)	0.6961(0.0194)
SPIN+INF	0.2195(0.0046)	0.2106(0.0422)	0.2551(0.0428)	0.1531(0.0224)	0.4728(0.0194)	0.7629(0.1007)
ImputeFormer+INF	0.1776(0.0009)	0.1461(0.0013)	0.2085(0.0020)	0.1033(0.0070)	0.4197(0.0058)	0.6498(0.0046)

Method	GEF	ETTH1	ETTH2	ELECTRICITY	TRAFFIC	AIR
I.Original						
Mean	0.1750	0.0523	0.1797	0.1123	0.4359	0.1508
SAITS	0.1980(0.0092)	0.1027(0.0021)	0.2098(0.0125)	0.1176(0.0110)	0.4311(0.0151)	0.5006(0.0251)
BRITS	0.2021(0.0007)	0.1692(0.0105)	0.2384(0.0018)	0.1503(0.0003)	0.4535(0.0001)	0.6979(0.0086)
MRNN	0.2052(0.0001)	0.2184(0.0016)	0.2317(0.0001)	-	0.4540(0.0000)	0.7965(0.0018)
GPVAE	0.2087(0.0019)	0.1591(0.0072)	0.2365(0.0022)	0.1471(0.0001)	0.4465(0.0001)	0.6968(0.0044)
USGAN	0.2048(0.0023)	0.1549(0.0179)	0.2238(0.0085)	0.1447(0.0011)	0.4742(0.0048)	0.6840(0.0306)
SPIN	0.2120(0.0029)	0.2000(0.0509)	0.2414(0.0327)	0.1588(0.0113)	0.4609(0.0148)	0.6604(0.0802)
ImputeFormer	0.1820(0.0016)	0.1558(0.0033)	0.2125(0.0022)	0.1076(0.0012)	0.4249(0.0060)	0.6300(0.0119)
II.With Seg-4 Gain estimation						
Mean+SAITS	**0.1666(0.0007)**	**0.0522(0.0000)**	0.1796(0.0000)	**0.0972(0.0006)**	**0.4182(0.0022)**	**0.1490(0.0001)**
Mean+BRITS	0.1704(0.0000)	**0.0522(0.0000)**	0.1795(0.0000)	0.1078(0.0000)	0.4332(0.0000)	0.1507(0.0000)
Mean+MRNN	0.1707(0.0000)	0.0523(0.0000)	0.1795(0.0000)	-	0.4333(0.0000)	0.1508(0.0000)
Mean+GPVAE	0.1708(0.0000)	**0.0522(0.0000)**	0.1795(0.0000)	0.1069(0.0001)	0.4308(0.0004)	0.1507(0.0000)
Mean+USGAN	0.1704(0.0001)	**0.0522(0.0000)**	0.1795(0.0000)	0.1076(0.0002)	0.4251(0.0002)	0.1506(0.0000)
Mean+SPIN	0.1693(0.0013)	0.0523(0.0001)	0.1800(0.0003)	0.1047(0.0001)	0.4302(0.0010)	0.1502(0.0005)
Mean+ImputeFormer	**0.1666(0.0002)**	**0.0522(0.0000)**	**0.1794(0.0000)**	0.0991(0.0002)	0.4203(0.0014)	0.1498(0.0000)
III.With Seg-2 Gain estimation						
Mean+SAITS	0.1686(0.0005)	**0.0522(0.0000)**	0.1799(0.0001)	0.1003(0.0005)	0.4212(0.0017)	0.1491(0.0001)
Mean+BRITS	0.1724(0.0000)	**0.0522(0.0000)**	0.1795(0.0000)	0.1105(0.0000)	0.4355(0.0000)	0.1507(0.0000)
Mean+MRNN	0.1730(0.0000)	0.0523(0.0000)	0.1795(0.0000)	-	0.4356(0.0000)	0.1508(0.0000)
Mean+GPVAE	0.1730(0.0000)	**0.0522(0.0000)**	0.1795(0.0000)	0.1097(0.0001)	0.4335(0.0003)	0.1507(0.0000)
Mean+USGAN	0.1724(0.0001)	**0.0522(0.0000)**	0.1795(0.0000)	0.1098(0.0003)	0.4290(0.0001)	0.1506(0.0000)
Mean+SPIN	0.1733(0.0008)	0.0523(0.0000)	0.1803(0.0004)	0.1093(0.0003)	0.4343(0.0001)	0.1503(0.0005)
Mean+ImputeFormer	0.1688(0.0004)	0.0523(0.0001)	0.1795(0.0000)	0.1021(0.0002)	0.4231(0.0008)	0.1498(0.0000)

- **Combining different imputation methods can generally benefit the down stream forecasting.**
- **As the number of segments in the acceleration method decreases, the advantage slightly decreases, but it is still maintained.**

Future Work



➢ Other downstream work. Optimization?

Codes

https://github.com/hkuedl/Task-Oriented-Imputation

香 港 大 學
THE UNIVERSITY OF HONG KONG

Energy Digitalization Laboratory

Alibaba

Zhixian Wang

Linxiao Yang

Liang Sun

Qingsong Wen

Yi Wang