Towards Diverse and Coherent Augmentation for Time-Series Forecasting

Xiyuan Zhang, Ranak Roy Chowdhury, Jingbo Shang, Rajesh Gupta, Dezhi Hong

University of California, San Diego

xiyuanzh@ucsd.edu

Outline

- Motivation
- Method
- Experiments
- Conclusion

Outline

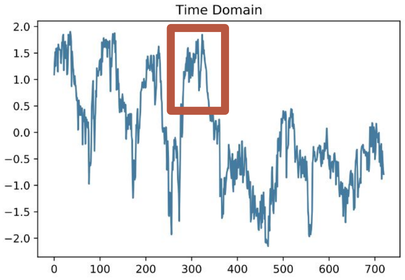
- Motivation
- Method
- Experiments
- Conclusion

Motivation

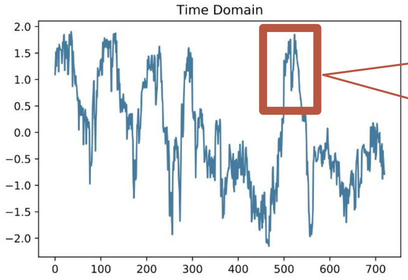
- Data scarcity in time series is very common
  - Malfunctioning sensors due to battery depletion
  - Privacy concerns in medical domain
  - Cold-start prediction for new stations or new stores

Motivation

- Data augmentation mitigates the data scarcity issue
- Existing time-series augmentation methods are mainly designed for classification



Original

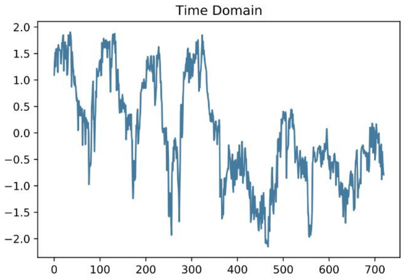


Permutation

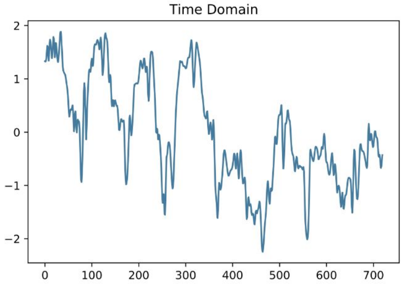
Class labels can be preserved even if augmentation changes the temporal dynamics

Motivation

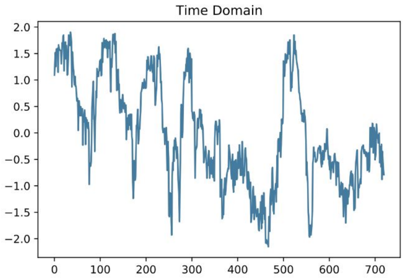
- Augmentation for time-series forecasting requires diversity as well as coherence with the temporal dynamics



Original



Filtering: not diverse



Permutation: not coherent

Motivation

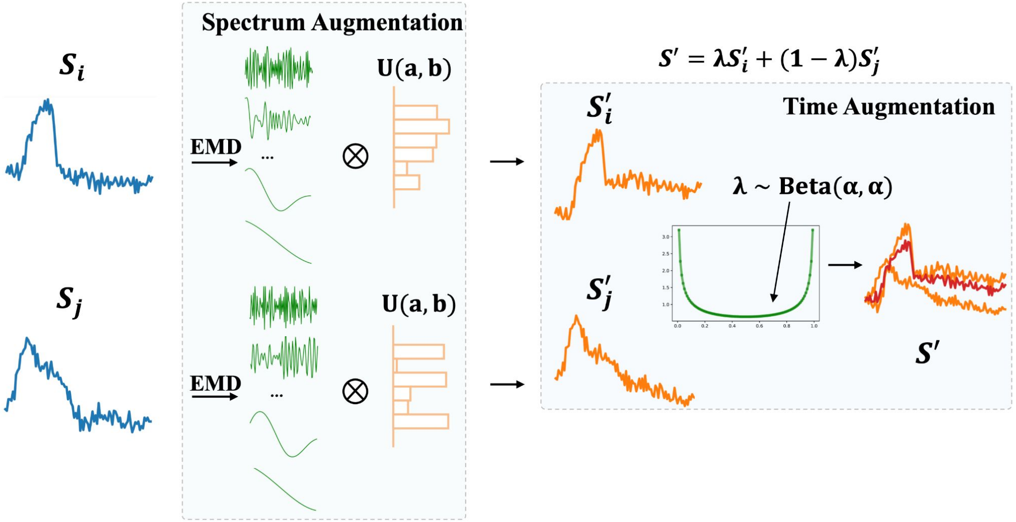
- Unlike images or text, time-series data generated by real-life physical processes have characteristics in both time and spectral domains
- Our contribution: Combining **S**pectral and **T**ime **Aug**mentation (**STAug**) to generate more diverse and coherent samples for time-series forecasting

Outline

- Motivation
- Method
- Experiments
- Conclusion

Method

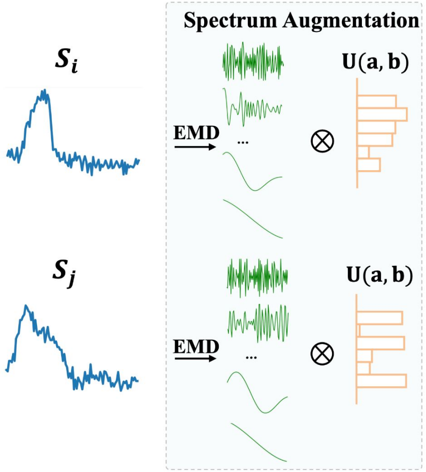
- Combine frequency-domain augmentation and time-domain augmentation



Method

- Frequency-domain augmentation: decompose time series with Empirical Mode Decomposition (EMD), and reassemble subcomponents with random weights

$$\mathbf{S} = \sum_{i=1}^{n} \mathrm{IMF}_i + \mathbf{R} \qquad \mathbf{S}' = \sum_{i=1}^{n} w_i \cdot \mathrm{IMF}_i$$

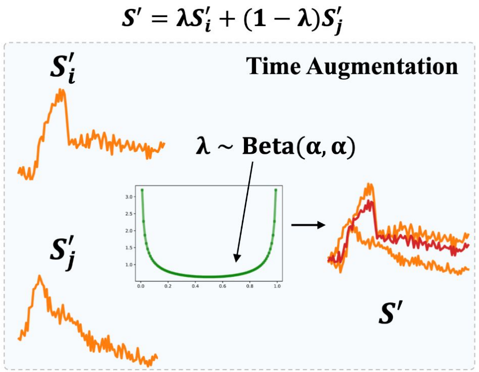


- Diversity: random weights to combine components
- Coherence: augmented samples have the same set of base components

Method

- Time-domain augmentation: mix up values at both history and future time steps

$$\mathbf{H}' = \lambda \mathbf{H_i} + (1-\lambda)\mathbf{H_j} \qquad \mathbf{F}' = \lambda \mathbf{F_i} + (1-\lambda)\mathbf{F_j}$$



- Diversity: random weights to combine series pairs
- Coherence: only generate linearly in-between samples by the interpolation nature of “mix up”

Method

- During training, we apply both frequency-domain and time-domain augmentation to obtain an augmented sample
- Optimize the downstream forecasting model with augmented samples as input

$$\mathcal{L} = \frac{1}{N} \sum_{i=1}^{N} ||\mathbf{Y_i} - \mathbf{F'_i}||_2^2$$

Outline

- Motivation
- Method
- Experiments
- Conclusion

Experiments

- Datasets: ETTh1, ETTh2, ETTm1, ETTm2, Exchange
- Forecasting backbone model: Informer (AAAI' 21 best paper)
- Baselines: basic random operations (WW, RobustTAD), decomposition-based augmentation (STL, EMD-R), pattern mixing method (DBA), generative method (GAN)

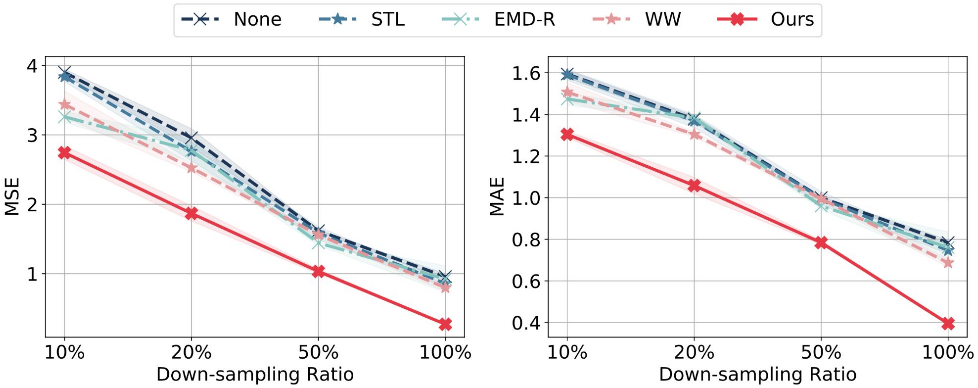
Experiments

- STAug reduces average MSE by 28.2% and average MAE by 18.0% compared with the best baseline for each of the five datasets

Methods		None		WW		RobustTAD		STL		EMD-R		GAN		DBA		STAug-noTime		STAug-noFreq		STAug	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	0.95±0.06	0.76±0.04	0.89±0.06	0.73±0.04	1.01±0.03	0.79±0.01	0.91±0.05	0.74±0.03	0.84±0.05	0.69±0.03	0.90±0.08	0.73±0.06	0.94±0.06	0.76±0.04	0.66±0.01	0.57±0.01	0.84±0.06	0.69±0.03	**0.65±0.01**	**0.57±0.01**
	192	0.98±0.02	0.77±0.02	1.00±0.12	0.78±0.08	0.95±0.02	0.74±0.01	0.94±0.02	0.74±0.02	0.94±0.04	0.73±0.04	0.99±0.04	0.76±0.03	0.98±0.02	0.77±0.02	**0.75±0.02**	**0.62±0.01**	0.96±0.00	0.76±0.01	0.77±0.03	0.63±0.02
	336	1.11±0.04	0.83±0.02	1.06±0.03	0.81±0.02	1.09±0.06	0.82±0.04	1.09±0.04	0.81±0.02	1.06±0.02	0.80±0.01	1.07±0.01	0.82±0.01	1.11±0.04	0.83±0.02	**0.86±0.02**	**0.68±0.01**	1.10±0.06	0.83±0.03	0.89±0.02	0.71±0.01
	720	1.18±0.03	0.86±0.02	1.20±0.02	0.88±0.01	1.15±0.02	0.84±0.01	1.19±0.02	0.86±0.01	1.19±0.02	0.87±0.01	1.22±0.04	0.89±0.01	1.18±0.03	0.87±0.02	0.99±0.06	0.75±0.03	1.16±0.01	0.86±0.01	**0.97±0.00**	**0.75±0.00**
ETTh2	96	3.08±0.62	1.38±0.13	3.40±0.48	1.46±0.11	3.03±0.25	1.38±0.05	3.04±0.48	1.37±0.11	3.44±0.29	1.45±0.05	2.91±0.36	1.38±0.09	3.01±0.54	1.37±0.12	**1.80±0.11**	**1.01±0.04**	2.98±0.57	1.37±0.14	1.87±0.30	1.05±0.09
	192	5.97±0.30	2.02±0.05	6.43±0.68	2.10±0.09	6.20±0.35	2.08±0.05	5.89±0.46	2.02±0.08	6.33±0.10	2.09±0.00	5.47±0.30	1.94±0.05	5.64±0.47	1.94±0.09	4.02±0.23	1.57±0.05	5.23±0.15	1.89±0.03	**3.47±0.24**	**1.48±0.06**
	336	4.78±0.33	1.83±0.08	4.97±0.07	1.85±0.05	5.29±0.15	1.88±0.01	4.77±0.24	1.83±0.05	5.44±0.18	1.95±0.05	4.91±0.31	1.87±0.06	4.77±0.32	1.83±0.08	3.39±0.27	**1.48±0.07**	4.37±0.26	1.74±0.05	**3.37±0.14**	1.51±0.04
	720	3.99±0.27	1.69±0.06	3.83±0.13	1.68±0.04	3.90±0.13	1.66±0.00	3.56±0.04	1.58±0.03	4.08±0.09	1.72±0.03	4.06±0.41	1.74±0.10	4.01±0.27	1.70±0.06	2.70±0.30	1.38±0.09	3.57±0.18	1.62±0.04	**2.62±0.17**	**1.34±0.08**
ETTm1	96	0.63±0.03	0.57±0.01	0.63±0.02	0.57±0.01	0.59±0.03	0.55±0.02	0.55±0.03	0.53±0.01	0.64±0.01	0.56±0.00	0.63±0.02	0.58±0.01	0.63±0.03	0.57±0.02	0.43±0.00	0.44±0.00	0.55±0.03	0.53±0.02	**0.40±0.00**	**0.42±0.00**
	192	0.80±0.04	0.67±0.02	0.79±0.07	0.67±0.03	0.80±0.08	0.66±0.04	0.74±0.06	0.65±0.03	0.68±0.01	0.59±0.01	0.84±0.06	0.69±0.02	0.80±0.03	0.68±0.02	0.54±0.02	0.51±0.00	0.67±0.04	0.59±0.02	**0.49±0.00**	**0.48±0.00**
	336	1.15±0.07	0.84±0.03	1.00±0.06	0.77±0.03	0.98±0.04	0.76±0.03	0.99±0.07	0.77±0.03	0.87±0.14	0.70±0.06	0.90±0.07	0.73±0.04	1.15±0.06	0.84±0.02	0.65±0.03	0.57±0.02	0.90±0.02	0.72±0.01	**0.63±0.03**	**0.56±0.01**
	720	1.13±0.04	0.81±0.00	1.08±0.07	0.79±0.02	1.06±0.04	0.79±0.03	1.18±0.02	0.82±0.00	1.01±0.10	0.75±0.05	1.23±0.12	0.86±0.04	1.13±0.04	0.81±0.01	0.83±0.03	0.66±0.01	0.93±0.10	0.72±0.04	**0.76±0.04**	**0.63±0.02**
ETTm2	96	0.42±0.06	0.51±0.04	0.39±0.02	0.48±0.02	0.39±0.02	0.48±0.01	0.39±0.03	0.49±0.03	0.40±0.08	0.48±0.05	0.33±0.03	0.43±0.03	0.42±0.05	0.51±0.04	0.30±0.03	0.39±0.01	0.33±0.06	0.42±0.04	**0.29±0.03**	**0.39±0.03**
	192	0.78±0.09	0.68±0.03	0.78±0.09	0.68±0.04	0.88±0.20	0.72±0.10	0.71±0.05	0.65±0.01	0.54±0.04	0.55±0.03	0.63±0.11	0.60±0.05	0.77±0.07	0.67±0.03	**0.42±0.00**	**0.49±0.01**	0.59±0.01	0.59±0.01	0.43±0.02	0.50±0.01
	336	1.52±0.07	0.95±0.02	1.43±0.08	0.92±0.02	1.39±0.18	0.91±0.07	1.40±0.06	0.91±0.03	1.29±0.04	0.88±0.02	1.45±0.12	0.93±0.05	1.57±0.10	0.96±0.03	0.94±0.19	0.74±0.07	1.29±0.11	0.88±0.03	**0.81±0.03**	**0.69±0.01**
	720	3.46±0.27	1.42±0.08	4.37±0.29	1.64±0.08	4.79±0.03	1.65±0.01	3.27±0.30	1.38±0.08	3.25±0.35	1.35±0.07	3.13±0.71	1.36±0.17	3.48±0.27	1.43±0.08	**2.65±0.47**	**1.20±0.10**	3.77±0.38	1.51±0.10	2.79±0.25	1.27±0.04
Exchange	96	0.96±0.04	0.78±0.01	0.80±0.06	0.69±0.02	0.96±0.04	0.79±0.02	0.86±0.05	0.75±0.02	0.92±0.19	0.76±0.08	0.94±0.06	0.77±0.03	0.96±0.04	0.78±0.01	0.29±0.01	0.40±0.00	0.81±0.02	0.71±0.01	**0.27±0.00**	**0.40±0.01**
	192	1.11±0.01	0.84±0.00	1.08±0.00	0.79±0.00	1.18±0.01	0.87±0.01	1.14±0.04	0.85±0.01	1.18±0.02	0.85±0.01	1.12±0.02	0.86±0.01	1.11±0.01	0.84±0.00	0.63±0.05	0.62±0.03	1.13±0.05	0.82±0.01	**0.61±0.04**	**0.61±0.02**
	336	1.61±0.04	1.01±0.02	1.70±0.07	1.00±0.02	1.56±0.03	1.01±0.00	1.55±0.02	0.99±0.01	1.61±0.09	1.00±0.04	1.47±0.11	0.97±0.03	1.60±0.04	1.01±0.02	1.01±0.13	0.81±0.05	1.44±0.02	0.94±0.01	**0.94±0.12**	**0.78±0.05**
	720	2.85±0.16	1.39±0.04	3.20±0.08	1.48±0.02	2.82±0.20	1.38±0.07	2.48±0.22	1.30±0.06	2.91±0.04	1.41±0.01	2.57±0.17	1.31±0.07	2.85±0.16	1.39±0.04	**1.41±0.24**	**0.93±0.08**	2.06±0.16	1.15±0.05	1.77±0.13	1.05±0.07

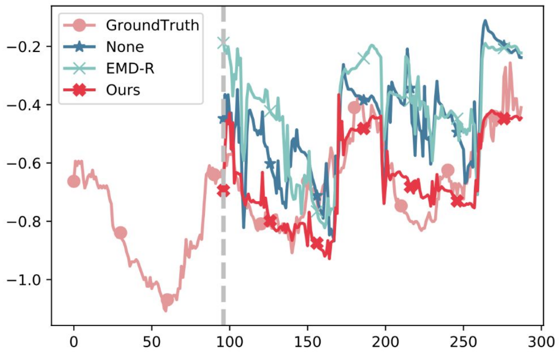
Experiments

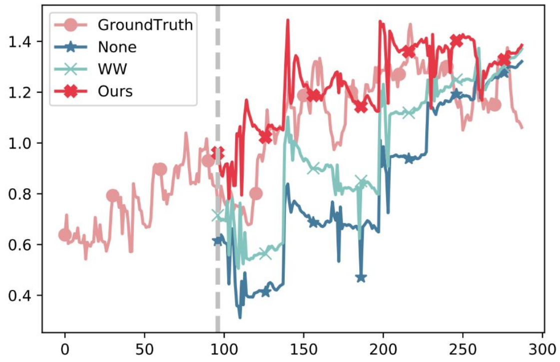
- STAug stays the most robust with respect to original sample size



Experiments

- STAug forecasts values that better align with the original time series





Outline

- Motivation
- Method
- Experiments
- Conclusion

Conclusion

- Diverse and coherent augmentation for time-series forecasting
  - In the frequency domain, the re-combined subcomponents of original time series are both diverse and preserve the original basic components
  - In the time domain, we adapt “mix up” to generate diverse and in-between coherent samples by linearly interpolating past and future parts of series
- Experiments on five datasets show that STAug best reduces the forecasting errors of the base model compared with existing augmentation methods.

Thanks!

Q & A