西湖大學
WESTLAKE UNIVERSITY

CIKM 2019 | BEIJING CHINA
November 03-07. 2019

Time Series Prediction with Interpretable Data Reconstruction

Qiangxing Tian, Jinxin Liu, Donglin Wang, Ao Tang

Presenter: Jinxin Liu

Machine Intelligence Laboratory, Department of Engineering,
Westlake University, Hangzhou, China

November, 2019

西湖大學 WESTLAKE UNIVERSITY

Outline

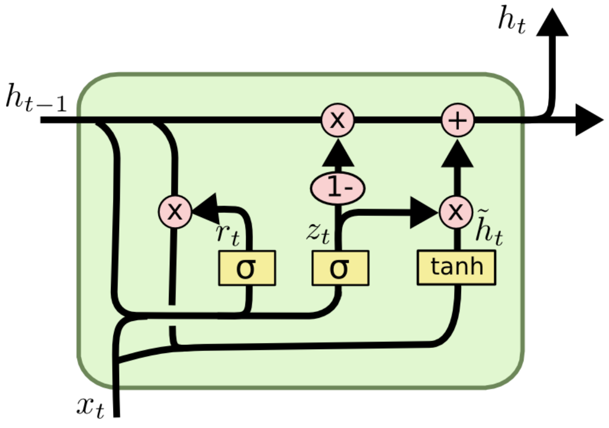
1. Background & Motivation
2. Methodology
3. Empirical Study

Outline

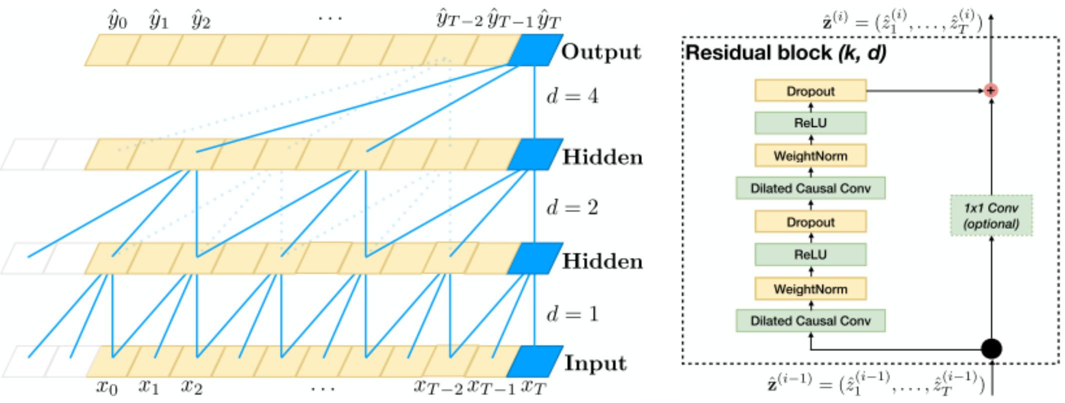
1. Background & Motivation
2. Methodology
3. Empirical Study

Background

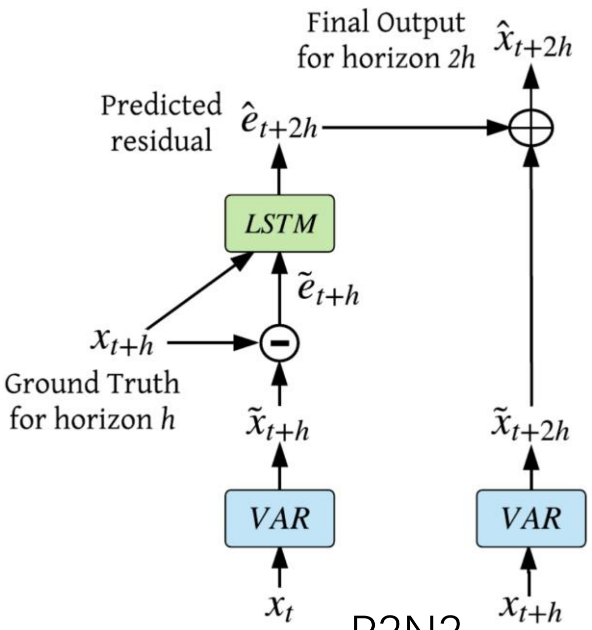
1. Statistical methods: ARIMA, VAR, SAR ⋯
2. Deep Neural Networks: RNN (LSTM), TCN[1] (CNN) ⋯
3. Hybrid methods: R2N2[2], LSTNet[3] ⋯



LSTM cell



TCN

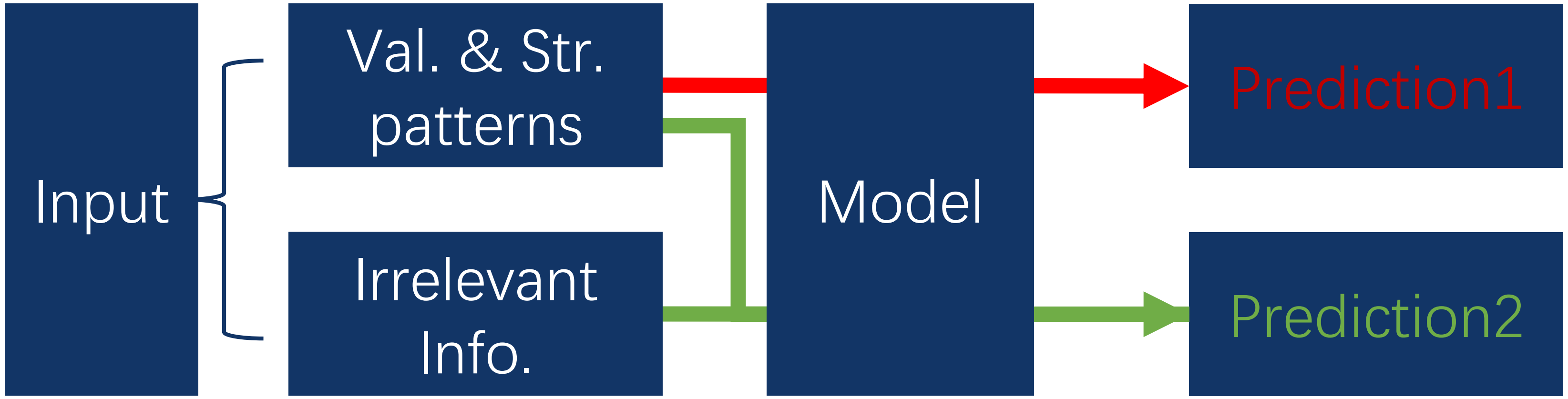


R2N2

[1] Bai, Shaojie, J. Zico Kolter, and Vladlen Koltun. "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling." arXiv preprint arXiv:1803.01271 (2018).
[2]Goel, Hardik, Igor Melnyk, and Arindam Banerjee. "R2N2: Residual recurrent neural networks for multivariate time series forecasting." arXiv preprint arXiv:1709.03159 (2017).
[3]Lai, Guokun, et al. "Modeling long-and short-term temporal patterns with deep neural networks." The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval. ACM, 2018.

Background

- Input = valuable & structural patterns + irrelevant information.
- Irrelevant information will lower the performance of model.
- How to extract the patterns which is task relevant without any prior?



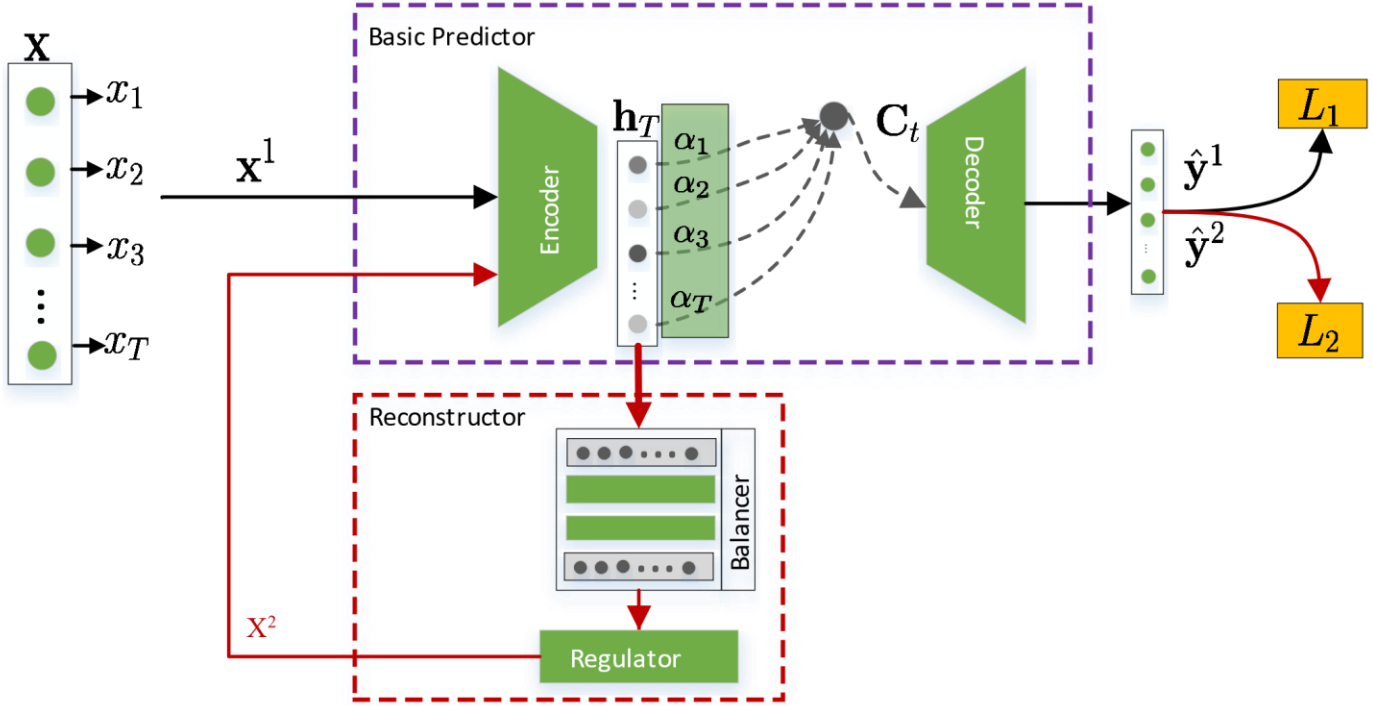
Outline

1. Background & Motivation
2. Methodology
3. Empirical Study

Methodology - overview

Framework

- Basic Predictor
  - Encoder
  - Decoder
- Reconstructor
  - Balancer
  - Regulator

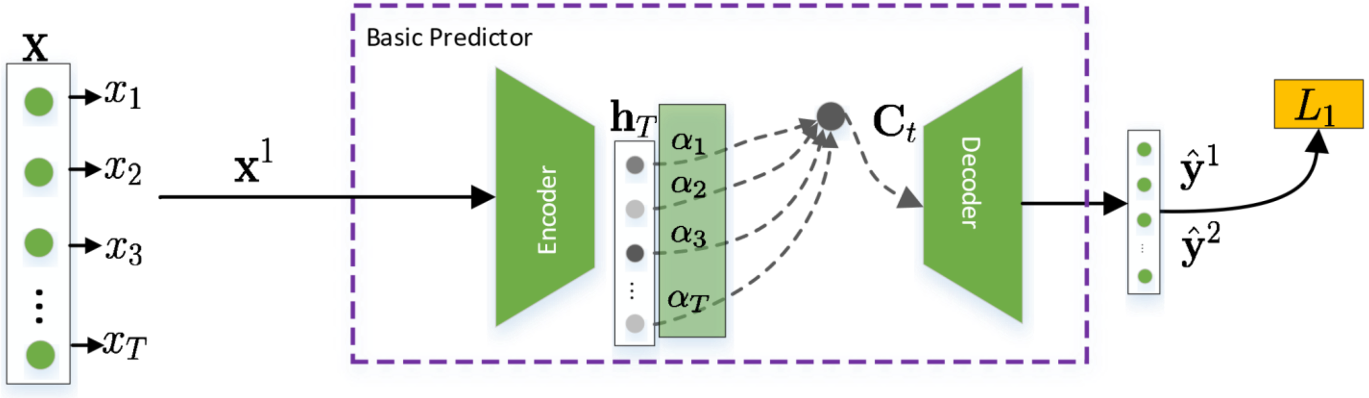


$$x \rightarrow x_1 \rightarrow h_T \rightarrow \hat{y}^1 \mapsto L_1$$
$$x \rightarrow x_1 \rightarrow h_T \rightarrow x_2 \rightarrow h_T \rightarrow \hat{y}^2 \mapsto L_2$$

西湖大學 WESTLAKE UNIVERSITY

Methodology – Stage1

- Basic Predictor
  - Encoder
  - Decoder

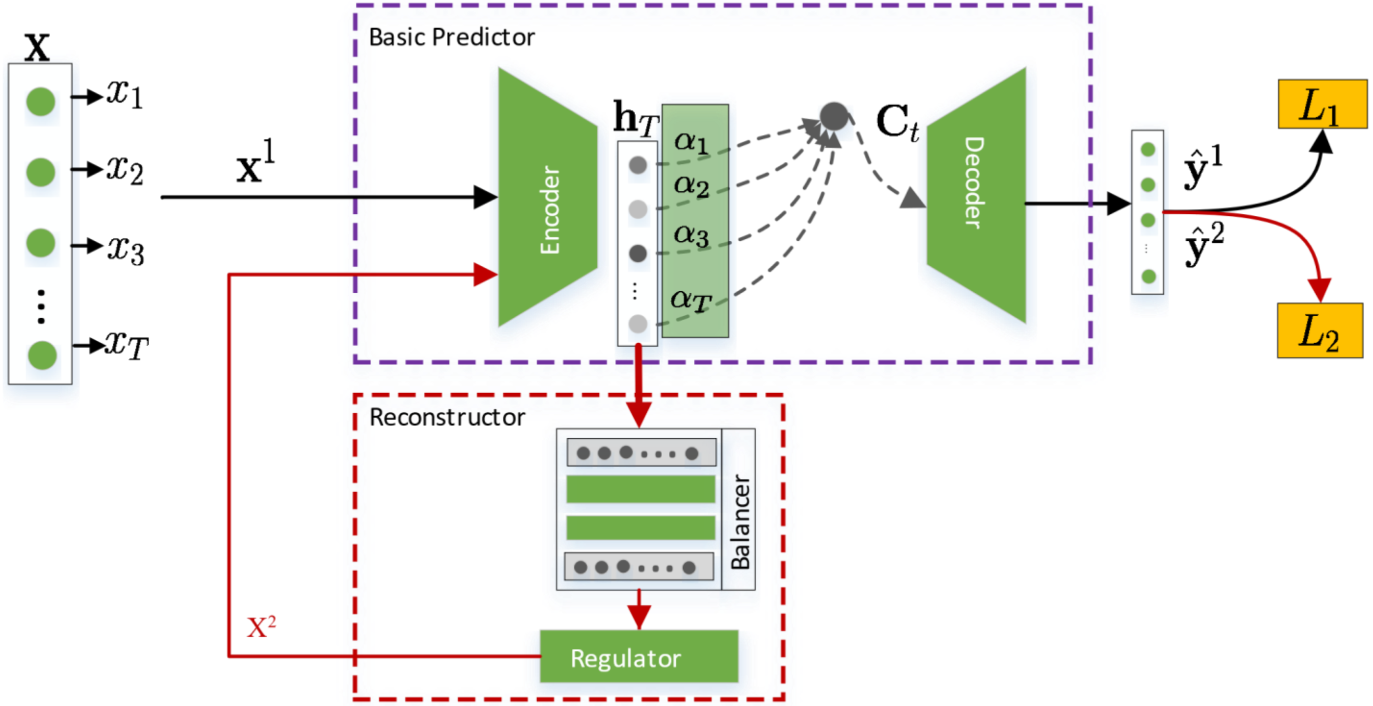


$$\mathrm{x} \rightarrow \mathrm{x}_1 \rightarrow \mathrm{h}_\mathrm{T} \rightarrow \hat{\mathrm{y}}^1 \mapsto \mathrm{L}_1$$

$$\mathrm{L}_1 = \mathrm{L}(y, \hat{y}^1)$$

Methodology – Stage2

- Basic Predictor
  - Encoder
- Reconstructor
  - Balancer
  - Regulator
- Basic Predictor
  - Encoder
  - Decoder

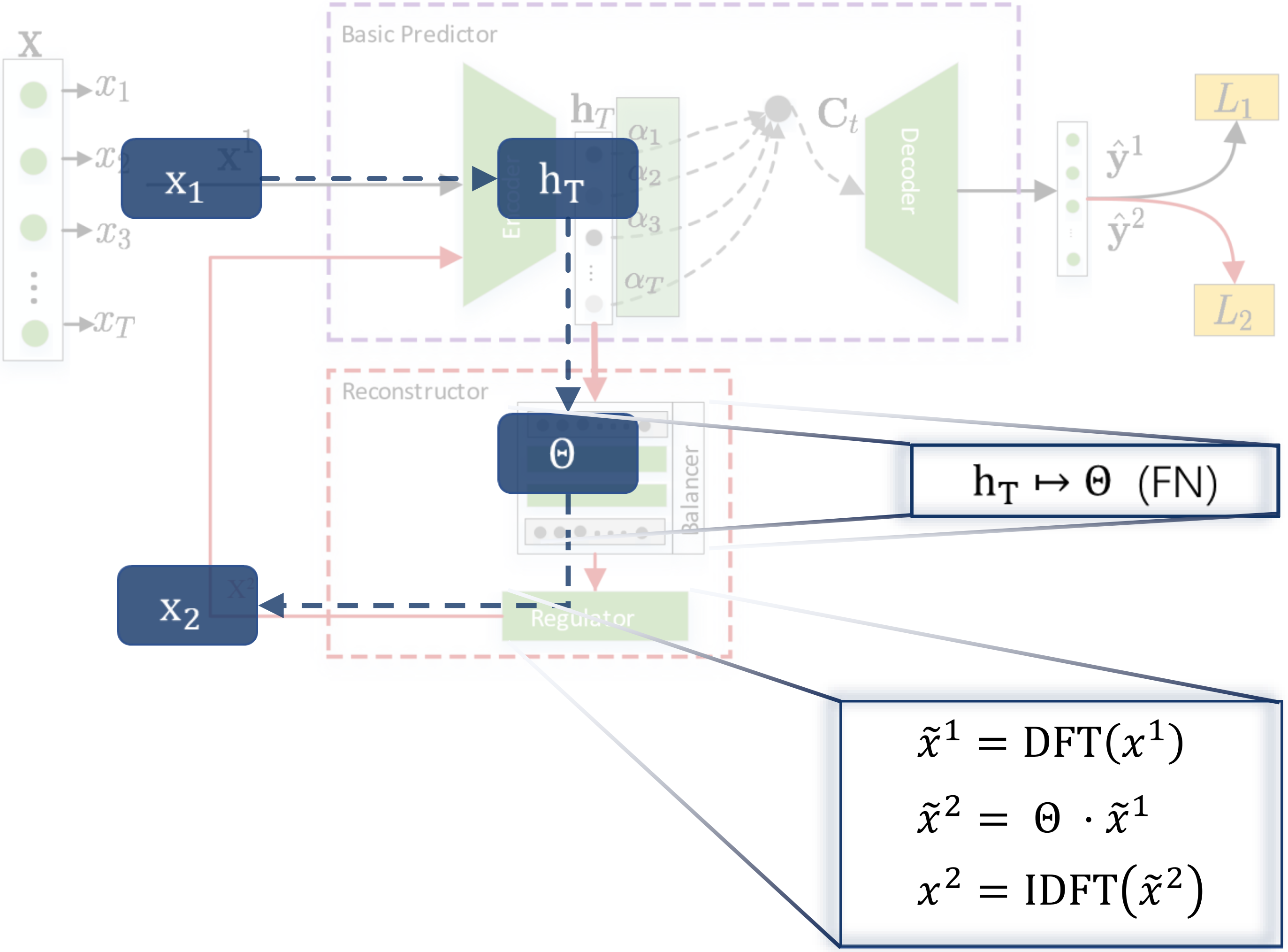


$$x \to x_1 \to h_T \to x_2 \to h_T \to \hat{y}^2 \mapsto L_2$$

$$L_2 = \gamma \cdot L\left(x^2, x^1\right) + (1-\gamma) \cdot L\left(y^2, \hat{y}^2\right)$$

Methodology - reconstructor

- Basic Predictor
  - Encoder
- Reconstructor
  - Balancer
  - Regulator
- Basic Predictor
  - Encoder
  - Decoder

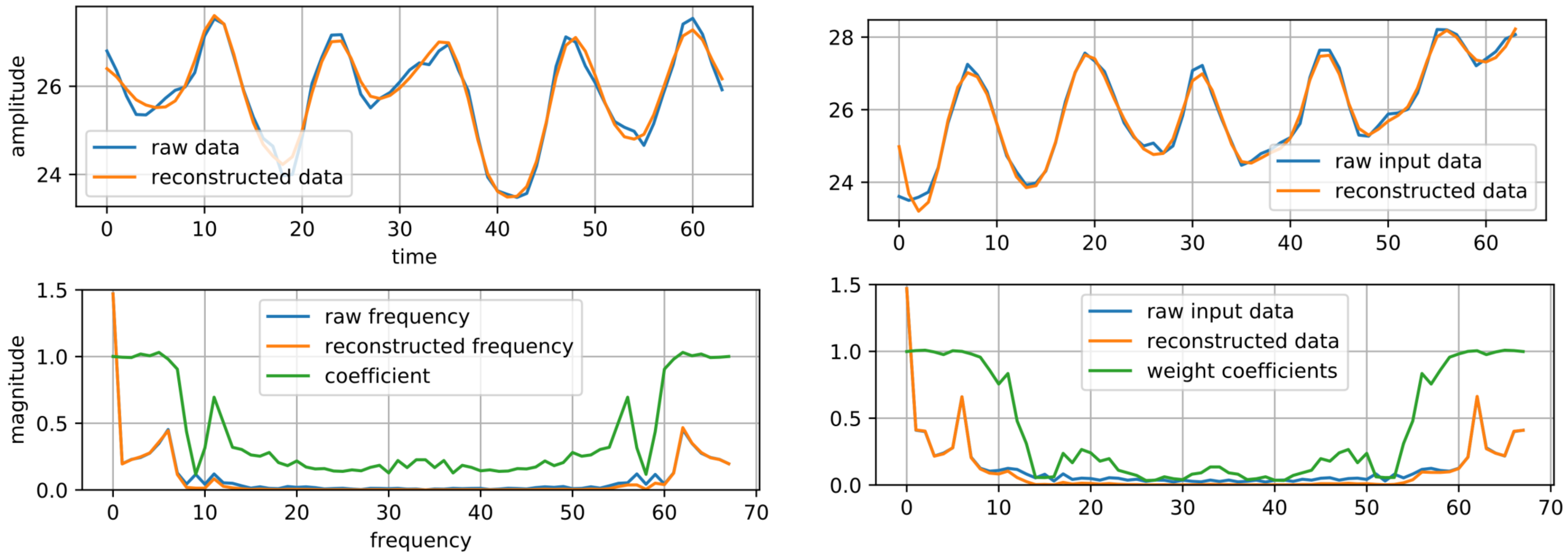


Outline

1. Background & Motivation
2. Methodology
3. Empirical Study

Empirical Study

Reconstruction: $x_1 \rightarrow h_T \rightarrow x_2$



西湖大學 WESTLAKE UNIVERSITY

Empirical Study

Overall performance: (IPR is our method)

Datasets	H	AR		Ridge		TCN		Wavelet-Trans		Seq2Seq		IPR	
		MSE	U	MSE	U	MSE	U	MSE	U	MSE	U	MSE	U
NINO 1-2	2	1.3667	0.0021	0.2782	0.0010	0.6004	0.0014	1.1718	0.0020	0.5800	0.0014	0.8918	0.0017
	4	3.4313	0.0034	0.4691	0.0012	0.6814	0.0015	0.9009	0.0017	1.1994	0.0020	**0.4477**	**0.0012**
	8	10.6932	0.0062	0.8739	0.0017	0.7875	0.0016	1.3262	0.0021	1.6922	0.0024	**0.5136**	**0.0013**
NINO 3	2	0.4010	0.0009	0.1776	0.0006	0.5530	0.0011	0.5341	0.0011	0.3218	0.0008	0.2082	0.0007
	4	1.8407	0.0020	0.3845	0.0009	0.3166	0.0008	0.6066	0.0011	0.7815	0.0013	**0.2862**	**0.0008**
	8	3.7377	0.0029	0.6229	0.0012	0.5916	0.0011	0.8459	0.0013	1.3398	0.0017	**0.1739**	**0.0006**
NINO 3-4	2	0.1394	0.0005	0.1247	0.0005	0.3578	0.0008	0.4345	0.0009	0.2429	0.0007	**0.1152**	**0.0005**
	4	0.6217	0.0011	0.2551	0.0007	0.2638	0.0007	0.5700	0.0010	0.3634	0.0008	**0.1526**	**0.0005**
	8	1.6354	0.0017	0.4561	0.0009	0.5871	0.0010	0.8947	0.0013	1.0797	0.0014	**0.0906**	**0.0004**
NINO 4	2	0.0483	0.0003	0.0567	0.0003	0.4068	0.0008	0.3309	0.0007	0.1151	0.0004	0.0739	0.0003
	4	0.1625	0.0005	0.1512	0.0005	0.1351	0.0004	0.3141	0.0007	0.2030	0.0006	**0.0633**	**0.0003**
	8	0.4922	0.0009	0.4209	0.0008	0.4999	0.0009	0.6071	0.0010	0.4306	0.0008	**0.0693**	**0.0003**
EP	2	19.5279	0.0029	4.3653	0.0014	4.1183	0.0013	5.2937	0.0015	6.6679	0.0017	6.0680	0.0016
	4	34.3566	0.0038	6.9597	0.0017	6.9498	0.0017	6.9967	0.0017	6.7628	0.0017	**6.6738**	**0.0017**
	8	70.5371	0.0060	7.7967	0.0019	7.2836	0.0017	8.9940	0.0020	15.0008	0.0026	**6.0237**	**0.0016**

Conclusion

- We propose a framework to learn the effective component without prior.
- Other learning algorithms or architectures are orthogonal to our framework and could be used to improve performance.
- Our approach learns the effective component in time series forecasting instead of artificially designing a filter via signal processing.
- Our model shows promising results and outperforms baselines.

Thank You!