Time Series Prediction with Interpretable Data Reconstruction

Qiangxing Tian[1,2*], Jinxin Liu[2*], Donglin Wang[2♀], Ao Tang[3]

[1] Zhejiang University; [2] Westlake University; [3] WeCar Technology Co., Ltd.
* Equal contribution; ♀ Corresponding author
{tianqiangxing, liujinxin, wangdonglin}@westlake.edu.cn

Introduction Background & Motivation

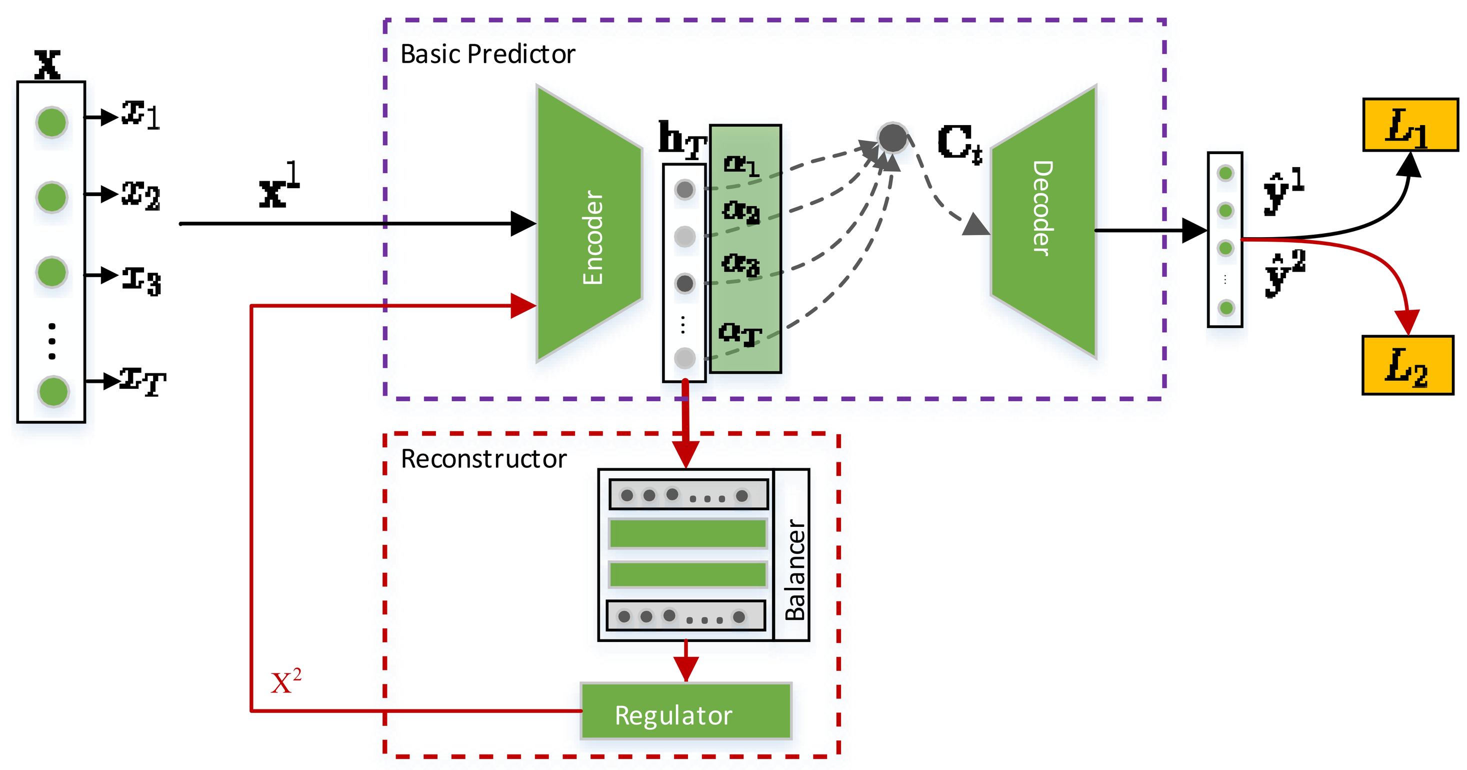
Time series prediction plays a key role in wide applications and has been investigated for a couple of decades. Nevertheless, most of the prior works fail to identify the most effective frequency components of time series before passing through the prediction, which induces the drop of the performance.

We consider that the input data contains valuable and structural pattern as well as irrelevant information. This irrelevant information will lower the performance of the model, so we need to extract the patterns which is task relevant without any prior.

Contributions

- To the best of our knowledge, this is the first work to learn the effective component in time series forecasting instead of artificially designing a filter via signal processing.
- Experiments results demonstrate the effectiveness in comparison to state-of-the-art baselines.

Model



First Stage: Initial training of the basic predictor. The encoder takes each x_t as input and updates the hidden state and cell state at each time stamp. The decoder aims to generate the basic prediction based on the attention mechanism.

Data flow: $x \to x^1 \to h_T \to \hat{y}^1 \mapsto L_1$

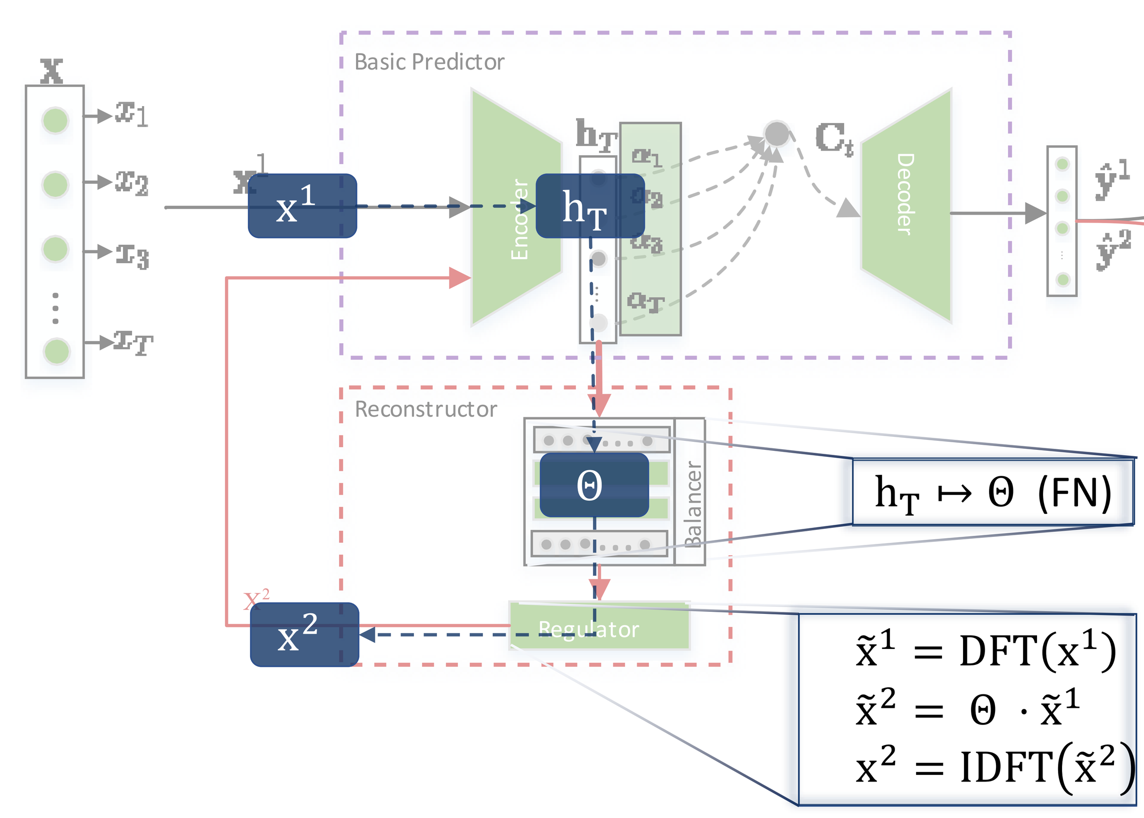
Loss function: $L_1 = L(y, \hat{y}^1)$

Second Stage: The original input data x^1 is reconstructed to x^2, and then x^2 is put into the basic predictor again and get the final prediction $\hat{y}^2$.

Data flow: $x \to x^1 \to h_T \to x^2 \to h_T \to \hat{y}^2 \mapsto L_2$

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

The **reconstructor** is composed of a balancer for coefficient calculation and a regulator for frequency adjustment.



Steps: We firstly feed the hidden state in the last time stamp in the encoder into the forward network (balancer). So we get the coefficient factor Θ. And we get the frequency spectrum of the input data $\tilde{x}^1$ by DFT. And multiply the coefficient factor Θ. So we get the reconstructed data x^2 by IDFT.

Data flow: $x^1 \begin{cases} x^1 \to h_T \to \Theta \\ x^1 \to \tilde{x}^1 \end{cases} \succ * \to \tilde{x}^2 \to x^2$

Experiment

Main Results

We choose two public datasets for experiments, ENSO (Nino Phenomenon) and EP (Electricity Price). The following table summarizes the evaluation results of various methods on all test data in terms of mean squared error (MSE) and Theil's U-statistics (U). As is shown in the below table, our method achieves the best performance compared to other 5 model in the majority experiments.

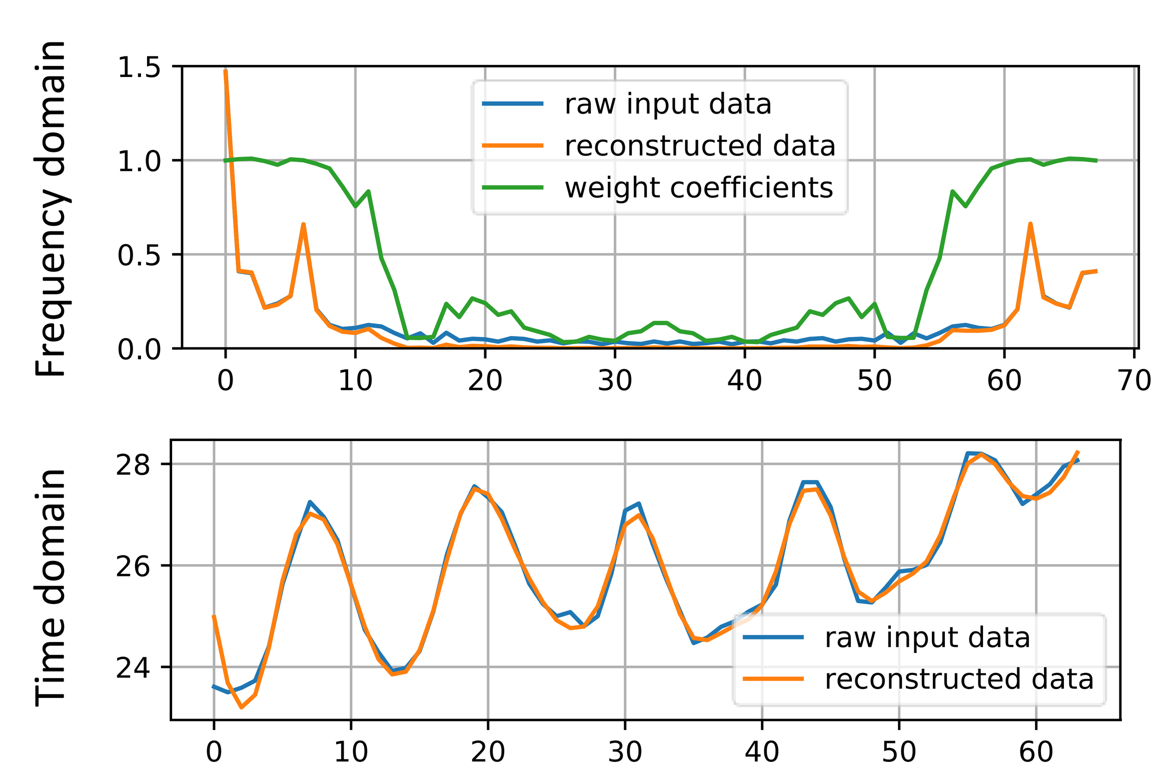
Datasets	H	AR		Ridge		TCN		Wavelet-T		Seq2Seq (BPSM)		IPR (Our method)	
		MSE	U	MSE	U	MSE	U	MSE	U	MSE	U	MSE	U
NINO 1-2	2	1.3667	0.0021	0.2782	0.0010	0.6004	0.0014	1.1718	0.0020	0.5800	0.0014	0.8918	0.0017
	4	3.4313	0.0034	0.4691	0.0012	0.6814	0.0015	0.9009	0.0017	1.1994	0.0020	**0.4477**	**0.0012**
	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**

Ablation Study

The core of our proposed IPR is the data reconstructor. After we remove the reconstructor, IPR degenerates back to BPSM. Obviously, BPSM is identical to Seq2Seq whose performance is illustrated in the upper table.

Interpretation

As illustrated in the below figures, it is observed that the effective frequency component can be extracted by our proposed reconstructor, where lots of the high frequency noise has been filtered out. In further, we can observe this change more clearly from the time-domain diagram where the reconstructed data is more smoother than the raw data.



Please note that other learning algorithms or architectures **are orthogonal to** our framework and could be used to improve performance. Anybody could design new basic predictor or reconsturctor, such as auto-encoder.

Conclusion

We have presented an interpretable data reconstructor for time series prediction in this paper. By integrating the data reconstructor and Seq2Seq model, the novel predictor is able to extract the most effective components of time series and thus exhibits an impressive performance in prediction compared to baselines.

西湖大學
WESTLAKE UNIVERSITY

CIKM2019