Two-Stage Framework for Seasonal Time Series Forecast

Qingyang Xu[1,2], Qingsong Wen[3], Liang Sun[3]

[1]MIT Operations Research Center, [2]MIT Laboratory for Financial Engineering

[3]Machine Intelligence Technology, DAMO Academy, Alibaba Group

Poster 2888 @ ICASSP 2021

Introduction

Forecasting univariate time series (TS) with seasonality has important real-world applications, such as proactive **auto-scaling of computing resources [1]**.

Self-supervised learning (SSL) enjoyed success in natural language modeling. However, it faces two main challenges in univariate time series forecast:

- Modeling univariate TS is a “small data” problem and the encoder network in SSL is very likely to overfit.
- Random masking in pre-training does not utilize seasonality property.

Two-Stage (2S) Framework

We propose two-stage model motivated by SSL. Stage 1 first predicts **future TS** (length $\boldsymbol{H}$) from **historical TS**

$$\hat{\mathbf{x}}_{fut} = f_1^*(\mathbf{x}_{his})$$

Stage 2 uses both **historical** and **future** TS (predicted by S1) to jointly predict the **forecast horizon** (length $\boldsymbol{h}$).

$$\hat{\mathbf{x}}_f = f_2^*(\mathbf{x}_{his}, \hat{\mathbf{x}}_{fut}) = f_2^*(\mathbf{x}_{his}, f_1^*(\mathbf{x}_{his}))$$

Two-Stage (2S) illustration

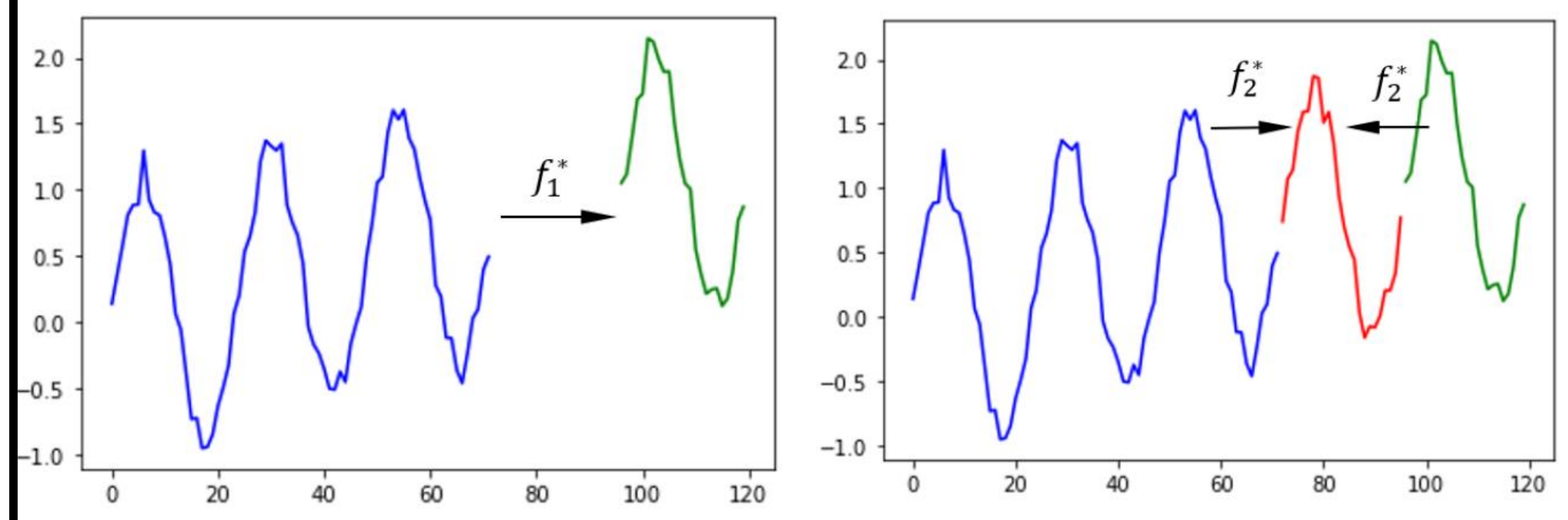

Figure 1. Illustration of 2S: Stage 1 (left) and Stage 2 (right).

Performance on M4 Hourly Data [2]

The M4 Competition Hourly dataset consists of 414 seasonal TS, each having between 700 to 960 points. We split each TS into two halves and use the first half for training and second for evaluation.

The two-stage model outperforms all baseline models in predicting the forecast horizon of length $\boldsymbol{h} = \mathbf{12}$ (Table 1). We use “MLP+MAR” model in both Stage 1 and Stage 2 predictions.

$$f^*(\mathbf{x}) = \mathbf{A}^{(0)}\mathbf{x} + \mathbf{b}^{(0)} + \mathrm{MLP}^{(m)}(\mathbf{x})$$

Model	MAPE	MAPE-95	RMSPE	RMSPE-95	RMSE	RMSE-95	MAE	MAE-95
Two-Stage	1.399	0.346	11.629	0.562	0.305	0.232	0.214	0.179
MLP+MAR	1.417	0.379	11.058	0.610	0.330	0.255	0.235	0.199
MLP	1.423	0.410	11.197	0.605	0.385	0.305	0.281	0.242
Deep-LSTM	1.539	0.459	11.857	0.652	0.422	0.341	0.320	0.278
MAR	1.551	0.416	12.275	0.672	0.349	0.275	0.253	0.216
RESTFul [3]	1.642	0.451	11.808	0.721	0.375	0.301	0.276	0.238
PrevPeriod	1.776	0.435	14.365	0.733	0.391	0.292	0.263	0.217
SPAR-h12	2.077	0.570	15.825	0.869	0.447	0.364	0.340	0.297

Table 1. Prediction performance on M4 Hourly data.

Two-stage Improves Baseline Models

We enhance the best baseline (BL) model with an additional Stage 1 model and improve its performance across different forecast horizon lengths h (Table 2).

h	Model	MAPE	MAPE-95	RMSPE	RMSPE-95	RMSE	RMSE-95	MAE	MAE-95
6	BL	1.265	0.332	10.008	0.542	0.285	0.219	0.201	0.169
	2S	1.214	0.305	10.102	0.496	0.266	0.201	0.185	0.155
12	BL	1.454	0.384	11.423	0.617	0.331	0.258	0.237	0.201
	2S	1.399	0.346	11.629	0.562	0.305	0.232	0.214	0.179
24	BL	1.511	0.405	11.833	0.651	0.349	0.273	0.251	0.214
	2S	1.489	0.374	11.900	0.614	0.319	0.247	0.226	0.191

Table 2. Including Stage 1 improves baseline models.

Optimize Future Horizon Length $\boldsymbol{H}^*$

A larger H incorporates more long-range TS structure but is also harder to predict. For forecast horizon $\boldsymbol{h} = 12$, the optimal performance is achieved for $\boldsymbol{H}^* = \mathbf{12}$.

H	MAPE	MAPE-95	RMSPE	RMSPE-95	RMSE	RMSE-95	MAE	MAE-95	S1 MSE
0	1.454	0.384	11.423	0.617	0.331	0.258	0.237	0.201	N/A
6	1.417	0.351	11.715	0.570	0.308	0.235	0.216	0.182	0.151
12	1.399	0.346	11.629	0.562	0.305	0.232	0.214	0.179	0.147
18	1.462	0.356	12.052	0.582	0.309	0.237	0.218	0.183	0.158
24	1.470	0.356	11.985	0.585	0.306	0.234	0.215	0.180	0.166

Table 3. Including Stage 1 improves baseline models.

Main References

[1] A Bauer, M Zufle, N Herbst, A Zehe, A Hotho, and S Kounev (2020), “Time Series Forecasting for Self-Aware Systems,” Proc. of the IEEE, 108(7) 1068–1093.

[2] Spyros Makridakis, Evangelos Spiliotis, and Vassilios Assimakopoulos (2020), “The M4 competition: 100,000 time series and 61 forecasting methods,” International Journal of Forecasting, 36(1) 54–74.

[3] Xian Wu, Baoxu Shi, Yuxiao Dong, Chao Huang, Louis Faust, and Nitesh V Chawla, “RESTFul: Resolution-aware forecasting of behavioral time series data,” in Proceedings of the 27th ACM International Conference on Information and Knowledge Management, 2018, pp. 1073–1082.