Refining Multivariate Forecasting with Adaptive Temporal-Contextual Learning

Jiecheng Lu [1] Xu Han [2] Shihao Yang [1]

[1]Georgia Institute of Technology [2]Amazon Web Services

Introduction

LTSF is vital but faces challenges in complex temporal relationships. Multivariate models often underperform due to inefficiencies in capturing series differences. We introduce ARM, a specialized multivariate LTSF architecture. ARM employs Adaptive Univariate Effect Learning (AUEL), Random Dropping(RD) training strategy, and Multi-kernel Local Smoothing (MKLS) to improve series pattern recognition and inter-series learning. It outperforms other models without added costs and enhances various LTSF architectures beyond the standard Transformer.

Problem: Capturing Correct Multivariate Temporal Dependencies

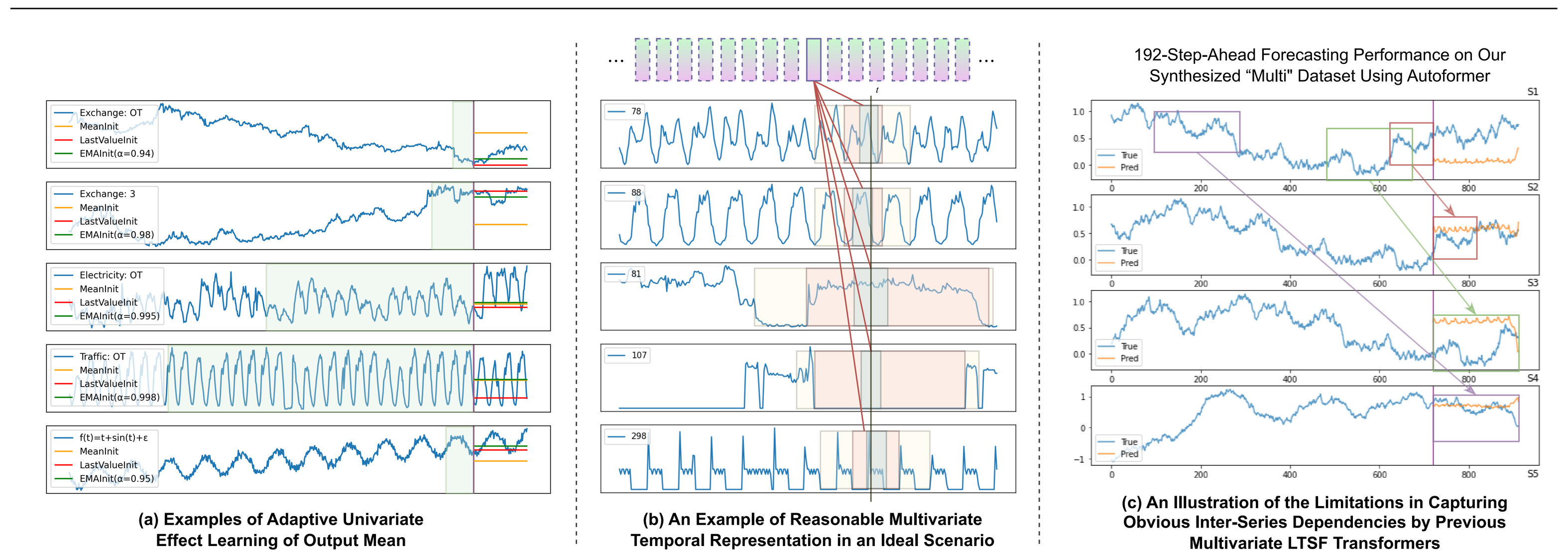

Figure 1. Three intuitive problems arising from wrongly handling input series with characteristic differences. (a) Adaptive Univariate Effect Learning of Output Mean. We illustrate how different methods estimate the output mean for five example time series. (b) The Necessity of Building Reasonable Multivariate Temporal Representation, which is demonstrated by five series from the Electricity dataset. (c) The Inability of Previous Models to Learn Obvious Inter-Series Dependencies. We illustrate the ineffectiveness of existing multivariate LTSF Transformers using a dataset "Multi", where the subsequent three series are simple shifts of the first

Module A: Adaptive Univariate Effect Learning (AUEL)

AUEL improves long-term and short-term forecasting in multivariate models. It adjusts outputs with EMA and uses MoE for better temporal pattern and inter-series understanding. Through preprocessing and inverse processing, AUEL ensures accurate forecasting, capturing time patterns and series relationships effectively.

Module R: Random Dropping and Inter-series Dependency Learning (RD)

Random Dropping strategy boosts multivariate models by understanding inter-series links. By randomly zeroing certain series during training, it helps models learn subset contributions to forecasting, reducing overfitting. This model ensemble create submodels for different series relationships and selecting the best ones.

Module M: Multi-kernel Local Smoothing (MKLS) block

MKLS tackles challenges in multivariate Transformer by enhancing temporal representations and locality. MKLS employs multiple 1D conv layers and channel-wise attention. It has two functions: Pre-MKLS for building local multivariate input representations, and Post-MKLS as adjustable local attention. MKLS enhances understanding of multivariate local temporal structures.

Overall Architecture

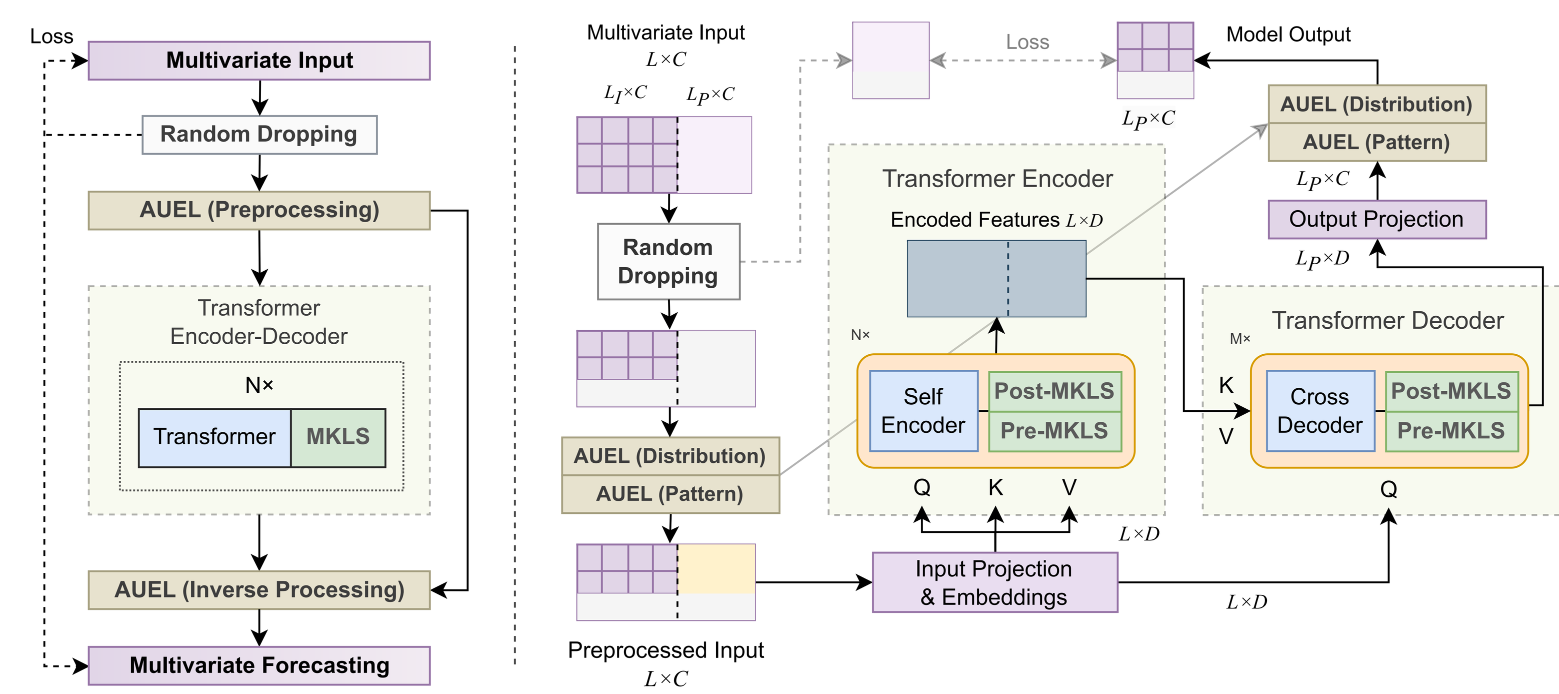

Figure 2. Overall Architecture of ARM with Vanilla Transformer as encoder-decoder. The left side depicts the global workflow incorporating the AUEL, Random Dropping, and MKLS modules. The right side illustrates the specific computational process employed by the model when training on a given multivariate time series input.

Structure of Module A (AUEL)

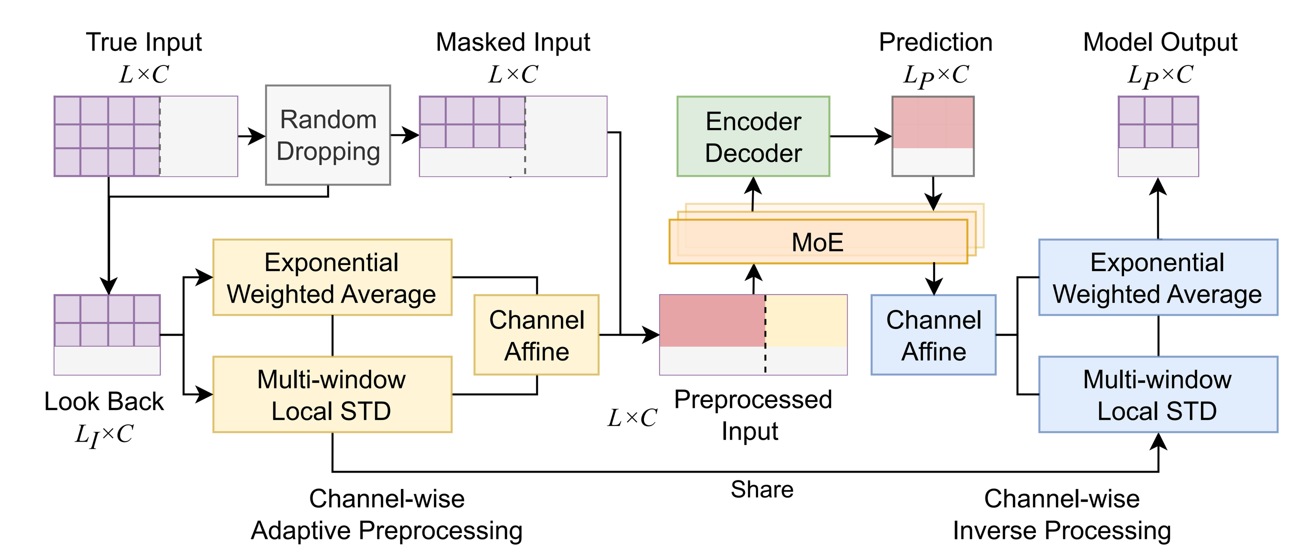

Figure 3. Structure of Adaptive Univariate Effect Learning (AUEL)

Structure of Module M (MKLS)

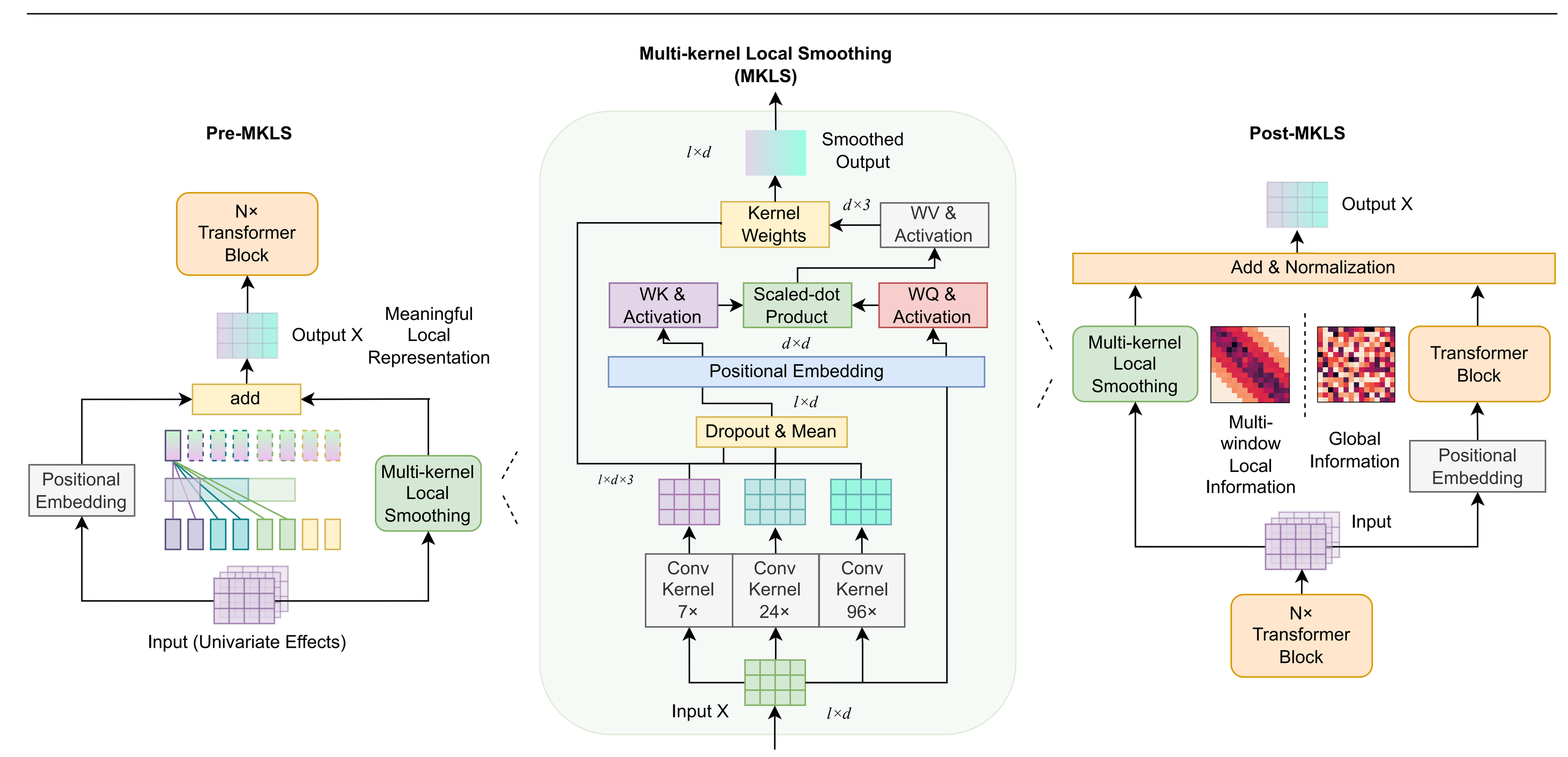

Figure 4. Structure of Multi-Kernel Local Smoothing (MKLS)

Effects of A/R/M Modules

1. (± ARM)		Vanilla		Vanilla+ARM		Autoformer		Autoformer+ARM		Informer		Informer+ARM	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Electricity	96	0.360	0.434	**0.125**	**0.222**	0.337	0.423	0.132	0.231	0.922	0.791	0.137	0.238
ETTm1	96	0.917	0.710	**0.287**	**0.340**	0.544	0.497	0.288	0.346	0.848	0.666	0.293	0.347
1. (± A)		Vanilla		Vanilla+ARM		Autoformer		Autoformer+A		Informer		Informer+A	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Electricity	96	0.360	0.434	**0.130**	**0.228**	0.337	0.423	0.131	0.231	0.922	0.791	0.139	0.242
ETTm1	96	0.917	0.710	**0.298**	0.357	0.544	0.497	0.304	0.355	0.848	0.666	0.302	**0.354**
3. (± R)		Vanilla		Vanilla+R		Autoformer		Autoformer+R		Informer		Informer+R	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Electricity	96	0.360	0.434	0.341	0.416	0.337	0.423	**0.207**	**0.320**	0.922	0.791	0.402	0.483
ETTm1	96	0.917	0.710	0.744	0.611	0.544	0.497	**0.431**	**0.458**	0.848	0.666	0.458	0.460
3. (± M)		Vanilla		Vanilla+M		Autoformer		Autoformer+M		Informer		Informer+M	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Electricity	96	0.360	0.434	0.378	0.453	0.337	0.423	**0.221**	**0.332**	0.922	0.791	0.832	0.724
ETTm1	96	0.917	0.710	0.834	0.649	**0.544**	0.497	0.563	**0.489**	0.848	0.666	0.809	0.623
5. (± RM)		Vanilla		Vanilla+RM		Autoformer		Autoformer+RM		Informer		Informer+RM	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Electricity	96	0.360	0.434	0.296	0.384	0.337	0.423	**0.200**	**0.315**	0.922	0.791	0.272	0.386
ETTm1	96	0.917	0.710	0.713	0.579	0.544	0.497	**0.272**	**0.356**	0.848	0.666	0.397	0.454

Conclusion

ARM addresses the challenge of handling time series with significant characteristic differences by extracting univariate effects, building multivariate representations, and identifying effective series combinations for forecasting. With only a minor increase in computational requirements, ARM enables a vanilla Transformer to achieve state-of-the-art performance in LTSF tasks. Its modular design allows easy integration into other LTSF models, highlighting its practical utility and potential for future research. ARM presents a promising approach for enhancing multivariate LTSF and offers opportunities for advancements in various time series domains.

References

[1] Shizhan Liu, Hang Yu, Cong Liao, Jianguo Li, Weiyao Lin, Alex X Liu, and Schahram Dustdar. Pyraformer: Low-complexity pyramidal attention for long-range time series modeling and forecasting. In *International conference on learning representations*, 2021.

[2] Yuqi Nie, Nam H. Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. A time series is worth 64 words: Long-term forecasting with transformers, 2023.

[3] Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting, 2022.

[4] Ailing Zeng, Muxi Chen, Lei Zhang, and Qiang Xu. Are transformers effective for time series forecasting?, 2022.

[5] Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 11106–11115, 2021.

[6] Tian Zhou, Ziqing Ma, Qingsong Wen, Xue Wang, Liang Sun, and Rong Jin. Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting, 2022.

 jlu414@gatech.edu