The Fortieth AAAI Conference on Artificial Intelligence (AAAI-26)

CompRestacking: Capturing Channel Dependency in Highly Correlated Multivariate Time Series Data (Student Abstract)

Min Kim, Ohyun Jo*

Chungbuk National University, 28644, Cheongju, Republic of Korea
kmin010287@chungbuk.ac.kr, ohyunjo@chungbuk.ac.kr

Abstract

The consideration of channel correlation is crucial for improving the performance of multivariate time series forecasting. However, existing models fail to capture it in homogeneous and highly correlated channels. In this work, we introduce CompRestacking (Compression Restacking), a strikingly intuitive and effective method to address this problem. The approach consists of three main components: (1) PCC-Restacking for correlation-aware channel ordering, (2) Temporal embedding for time encoding, and (3) Aggregation compression for compact token generation. CompRestacking consistently outperforms in experiment results. The results demonstrate that CompRestacking leverages strong channel correlations for improved performance.

Introduction

Multivariate time series (MTS) data are widely used in real-world domains, including health care (Morid, Sheng, and Dunbar 2023), communication (Chen et al. 2025), electricity (Iftikhar et al. 2024), energy consumption (Kardakos et al. 2013), and transportation (Kang and Jo 2024). In these environments, accurate and robust forecasting based on observed data is essential for decision making, policy formulation, and strategic planning (Han et al. 2024). In particular, channel correlation is crucial to improve the performance of MTS forecasting (Chen et al. 2024; Han, Ye, and Zhan 2024).

To investigate how channel dependencies influence prediction performance, pilot experiment conduct a systematic analysis employing widely adopted datasets and three forecasting benchmark models. The datasets in this experiment exhibit distinct characteristics: ETT (Zhou et al. 2021), with heterogeneous channel properties; ECL (Li et al. 2019), with homogeneous but weak correlations; and Seoul Traffic (Kang and Jo 2025), with both homogeneous and strongly correlated channels.

As demonstrated in Table 1, the amount of error is markedly elevated in the Seoul Traffic dataset. This striking divergence underscores the inherent limitations of existing correlation learning methodologies. Therefore, it is evident that a methodological advancement capable of more effectively accounting for channel correlations is required to play a decisive role in enhancing forecasting performance. In this study, we introduce ***CompRestacking*** (Compression Restacking), an intuitive and powerful framework specifically designed to structurally capture channel correlations. To this end, we particularly emphasize validating its effectiveness on the Seoul Traffic dataset, thereby demonstrating its capability in modeling strong channel dependencies.

Dataset	Model	MSE	MAE
ETTh1	TimesNet	0.0779	0.2099
	TimesMixer	0.0729	0.2048
	iTransformer	0.0708	0.2038
ETTh2	TimesNet	0.1914	0.3395
	TimesMixer	0.1707	0.3228
	iTransformer	0.1793	0.3332
ECL	TimesNet	0.0880	0.2266
	TimesMixer	0.0518	0.1609
	iTransformer	0.0643	0.1881
Seoul Traffic	TimesNet	**0.3100**	**0.3872**
	TimesMixer	**0.3349**	**0.4312**
	iTransformer	**0.4024**	**0.4893**

Table 1: Prediction errors increase significantly in homogeneous and highly correlated datasets.

Model	Metric	Horizon			
		96	120	144	168
Ours	MSE	**0.2898**	**0.2268**	**0.2266**	**0.2436**
	MAE	**0.3859**	**0.3138**	**0.3110**	**0.3330**
TimesNet	MSE	0.3154	0.3328	0.3081	0.3100
	MAE	0.4017	0.4165	0.3969	0.3872
TimeMixer	MSE	0.3504	0.3333	0.3289	0.3349
	MAE	0.4290	0.4294	0.4261	0.4312
iTransformer	MSE	0.4024	0.4064	0.4109	0.4024
	MAE	0.4893	0.4931	0.4973	0.4893
FEDformer	MSE	0.3629	0.3650	0.3713	0.3737
	MAE	0.4510	0.4521	0.4582	0.4614

Table 2: Forecasting results measured by MSE and MAE for different prediction horizons.

CompRestacking

PCC-Restacking. PCC-Restacking is a channel ordering methodology for MTS based on the Pearson Correlation Coefficient (PCC) (Benesty et al. 2009). As shown in Figure 1-

---

*Corresponding author: Ohyun Jo

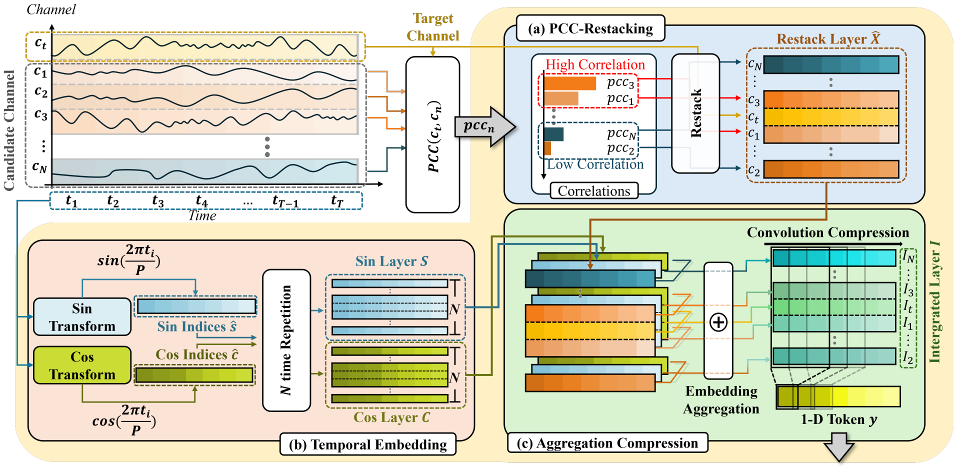

Figure 1: Overview of CompRestacking

(a), the procedure begins by positioning the target channel c_t at the center of the input series. We compute the PCC between c_t and each candidate channel c_n. In $Restack(\cdot)$, the channel with the highest correlation is then stacked either directly above or below the current stack according to pcc_n. At this stage, the target channel is redefined as the set of stacked channels. This procedure is applied iteratively until all N channels are arranged into restack layer $\hat{X}$.

$$pcc_n = PCC(c_t, c_n), \quad \hat{X} = Restack(c_t, \{pcc_n\}_{n=1}^{N}). \tag{1}$$

Temporal Embedding. Temporal information t_i is extracted from the original series. This approach improves predictive accuracy by accounting for data periodicity. The temporal granularity of the time embedding can be determined flexibly according to daily, weekly, monthly, or seasonal characteristics. As shown in Figure 1-(b), we transform time index $i_{i=1}^{T}$ using sinusoidal functions with period P:

$$e_i^s = \sin(2\pi t_i P)\,, \quad e_i^c = \cos(2\pi t_i P)\,. \tag{2}$$

where e_i^s and e_i^c represents the elements of sin indices $\hat{s}$ and cos indices $\hat{c}$ respectively. $\hat{s}$ and $\hat{c}$, following the aforementioned process, are repeated as N times to transform indices into Sin layer S and Cos layer C.

Aggregation Compression. Aggregation compression is adopted to condense essential information into a 1-dimensional token y. As shown in Figure 1-(c), Restack layer $\hat{X}$, Sin layer S, and Cos layer C are aggregated to make the integrated series $I = \hat{X} + S + C$. Subsequently, a convolution layer is employed to compress I via convolution along the time axis into a 1-dimensional token $y \in R^{K \times 1}$, where K denotes the length of the token.

Experiments

Experimental Setup. For performance evaluation of CompRestacking, experiments were conducted on the Seoul Traffic data set, comprising measurements from 105 sections of the Seoul highway with less than 1% missing values. The experiments focus on predicting future traffic speeds for a specific target road section.

Four channel-dependent benchmark models are compared with the proposed method in this experiment: TimesNet (Wu et al. 2022), CNN-based; TimeMixer (Wang et al. 2024), MLP-based; iTransformer (Liu et al. 2023) and FEDformer (Zhou et al. 2022), Transformer-based. All benchmark models follow the same lookback window $L = \{24, 48, 72, 96\}$ and forecasting horizon $H = \{96, 120, 144, 168\}$.

Experimental Results. Table 2 summarizes performance across horizons. CompRestacking consistently achieved the best performance across all forecasting horizons. Compared to CNN-based TimesNet, the proposed method achieves a 22.07% reduction in terms of MSE and 16.14% reduction in terms of MAE. Compared to the Transformer-based iTransformer, the proposed method achieves a reduction of 39.17% in terms of MSE and a reduction of 31.76% in terms of MAE. Compared to FEDFormer, CompRestacking reduces the amounts of error by 33% and 26.28% in terms of MSE and MAE, respectively. Compared to TimeMixer, the reductions are 26.77% in terms of MSE and 21.68% in terms of MAE. The performance of CompRestacking demonstrates its ability to capture high inter-channel correlations under the same experimental environment.

A comprehensive analysis of all MSE and MAE results demonstrated improved performance for horizons longer than $H = 96$. In particular, $H = 144$ yielded the lowest forecasting error rates. Consequently, this study indicates CompRestacking achieves robust performance even at longer forecasting horizons.

Conclusion

The existing methodologies fail to capture channel dependencies in homogeneous and highly correlated channels. In this study, we presented CompRestacking, a strikingly intuitive and effective correlation capture method to address this problem. Experiments verified that CompRestacking consistently and robustly outperformed existing models in reducing forecasting errors.

Acknowledgements

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT). (No.2021R1A2C2095289) In part, this work was supported by the Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT). (No. 2021-0-00165, Development of 5G+ Intelligent Basestation Software Modem).

References

Benesty, J.; Chen, J.; Huang, Y.; and Cohen, I. 2009. Pearson correlation coefficient. In *Noise reduction in speech processing*, 1–4. Springer.

Chen, C.; Liu, X.; Zhao, S.; and Bilal, M. 2025. GAN-based solar radiation forecast optimization for satellite communication networks. *International Journal of Intelligent Networks*.

Chen, J.; Lenssen, J. E.; Feng, A.; Hu, W.; Fey, M.; Tassiulas, L.; Leskovec, J.; and Ying, R. 2024. From similarity to superiority: Channel clustering for time series forecasting. *Advances in Neural Information Processing Systems*, 37: 130635–130663.

Han, L.; Chen, X.-Y.; Ye, H.-J.; and Zhan, D.-C. 2024. Softs: Efficient multivariate time series forecasting with series-core fusion. *Advances in Neural Information Processing Systems*, 37: 64145–64175.

Han, L.; Ye, H.-J.; and Zhan, D.-C. 2024. The capacity and robustness trade-off: Revisiting the channel independent strategy for multivariate time series forecasting. *IEEE Transactions on Knowledge and Data Engineering*, 36(11): 7129–7142.

Iftikhar, H.; Gonzales, S. M.; Zywiołek, J.; and López-Gonzales, J. L. 2024. Electricity demand forecasting using a novel time series ensemble technique. *IEEE Access*, 12: 88963–88975.

Kang, S.; and Jo, O. 2025. Memory-Efficient Imagification for Light-weight Prediction Model of Multivariate Time-Series Data. *IEEE Access*.

Kang, S. W.; and Jo, O. 2024. Multivariate time-series imagification with time embedding in constrained environments (student abstract). In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 23535–23536.

Kardakos, E. G.; Alexiadis, M. C.; Vagropoulos, S. I.; Simoglou, C. K.; Biskas, P. N.; and Bakirtzis, A. G. 2013. Application of time series and artificial neural network models in short-term forecasting of PV power generation. In *2013 48th International Universities' Power Engineering Conference (UPEC)*, 1–6. IEEE.

Li, S.; Jin, X.; Xuan, Y.; Zhou, X.; Chen, W.; Wang, Y.-X.; and Yan, X. 2019. Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting. *Advances in neural information processing systems*, 32.

Liu, Y.; Hu, T.; Zhang, H.; Wu, H.; Wang, S.; Ma, L.; and Long, M. 2023. itransformer: Inverted transformers are effective for time series forecasting. *arXiv preprint arXiv:2310.06625*.

Morid, M. A.; Sheng, O. R. L.; and Dunbar, J. 2023. Time series prediction using deep learning methods in healthcare. *ACM Transactions on Management Information Systems*, 14(1): 1–29.

Wang, S.; Wu, H.; Shi, X.; Hu, T.; Luo, H.; Ma, L.; Zhang, J. Y.; and Zhou, J. 2024. Timemixer: Decomposable multiscale mixing for time series forecasting. *arXiv preprint arXiv:2405.14616*.

Wu, H.; Hu, T.; Liu, Y.; Zhou, H.; Wang, J.; and Long, M. 2022. Timesnet: Temporal 2d-variation modeling for general time series analysis. *arXiv preprint arXiv:2210.02186*.

Zhou, H.; Zhang, S.; Peng, J.; Zhang, S.; Li, J.; Xiong, H.; and Zhang, W. 2021. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, 11106–11115.

Zhou, T.; Ma, Z.; Wen, Q.; Wang, X.; Sun, L.; and Jin, R. 2022. Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting. In *International conference on machine learning*, 27268–27286. PMLR.