Knowledge Informed Time Series Forecasting

Zijie Pan
zijie.pan@uconn.edu
University of Connecticut
Storrs, Connecticut, USA

Dongjin Song
dongjin.song@uconn.edu
University of Connecticut
Storrs, Connecticut, USA

Abstract

Multivariate Time Series (MTS) forecasting often faces challenges due to evolving dependencies and catastrophic forgetting across different regimes. To address these challenges, we propose Knowledge-Informed Time Series Forecasting, a novel framework that leverages structural knowledge to capture and maintain regime-specific dependencies, incorporating an innovative representation-matching scheme to select representative multivariate time series samples for memory replay. Experiments on synthetic and real-world datasets demonstrate KI-TSF's superior performance in forecasting accuracy and dependency inference over state-of-the-art methods.

Keywords

Time Series, Continual Learning, Graph Learning

ACM Reference Format:
Zijie Pan and Dongjin Song. 2025. Knowledge Informed Time Series Forecasting. In *Proceedings of 31th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD'25).* ACM, New York, NY, USA, 3 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

Multivariate time series (MTS) forecasting aims to predict future samples from multiple time series based on their historical values and has shown its importance in various applications [5, 6]. Despite the promising forecasting accuracy, MTS forecasters can only handle one regime (i.e., a period that exhibits consistent statistical properties or underlying dynamics) of MTS data. In real-world applications, different regimes of MTS data are often continuously collected under different operational logic of the target system. In this scenario, where regimes arrive sequentially, the major challenge for MTS forecasting models is to keep track of the latest regime while maintaining forecasting performance on the previous ones. While an intuitive and efficient solution is to retrain the forecaster periodically over the newly collected regime, this will inevitably lead to the catastrophic forgetting issue, *i.e.*, the learned dependency structures cannot be maintained over existing regimes and the forecasting performance will deteriorate accordingly, as shown in Figure 1.

We address this challenge by introducing the Knowledge-Informed Time Series Forecasting (KI-TSF) framework, which sequentially learns and retains meaningful dependency structures across different regimes for multivariate time series forecasting. Additionally, we propose a novel representation-matching memory replay strategy, selecting MTS samples that maximize temporal coverage and effectively preserve the underlying temporal dynamics and dependency structures, thereby enhancing continual forecasting performance.

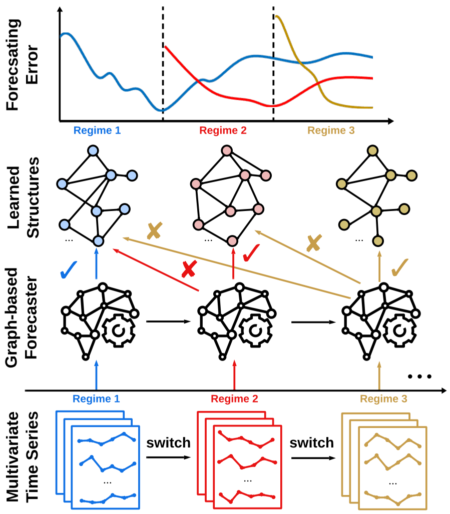

Figure 1: An illustration depicting the catastrophic forgetting of learned dependency structures

2 Structural Knowledge Informed Forecasting Model

Dynamic Graph Inference Module Different from the existing works that generate a static graph at the regime level [1, 17], we aim to model the variable dependencies of MTS as a dynamic graph at the granularity of an input window. Following [12], we explicitly model and parameterize each edge for all node pairs. For a possible edge connecting node i and j, we use a temporal encoding function as a feature extractor to yield node embedding z_i and z_j (*i.e.*, $z_* = \Phi(X_{*,t-\tau:t-1})$) which are concatenated as the edge embedding. Next, we use another generic mapping to finalize the edge generation as $\hat{A}_{ij} = \Psi(z_i \| z_j)$. The output graph $\hat{A} \in \mathbb{R}^{N \times N}$ summarizes the variable interactions from the temporal dynamics within a sequence, and can be further used to generate the forecasts. Note that there are multiple choices to parameterize Φ and Ψ, where we use stacked convolution layers and a multilayer perceptron (MLP), respectively.

To further exploit the structural and temporal dependencies and produce forecasts, we design a graph-based forecasting module consisting of multiple Temporal Graph Convolution (TGConv) blocks, which leverage a dilated causal convolution [2, 15] to effectively capture forward dynamics of time series.

KDD'25, Toronto, Canada

ACM ISBN 978-1-4503-XXXX-X/2018/06
https://doi.org/XXXXXXX.XXXXXXX

Table 1: Experiment Results for 12 Horizon Prediction. (Lower MAE and RMSE for AP mean better; When AP is comparable, lower MAE and RMSE for AF mean better.)

Model	Traffic-CL				Solar-CL				HAR-CL ($\times10^{-2}$)				Synthetic-CL($\times10^{-2}$)			
	AP ↓		AF		AP ↓		AF		AP ↓		AF		AP ↓		AF	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
VAR$_{\text{seq}}$	88.19	126.01	58.38	80.58	167.30	534.42	205.27	658.80	19.59	28.38	1.93	2.19	22.34	32.70	9.18	13.54
ARIMA$_{\text{seq}}$	141.75	159.89	77.61	77.40	14.97	18.92	4.75	2.92	40.68	52.87	2.35	2.38	42.24	43.51	13.39	12.98
STGCN$_{\text{seq}}$	30.45	49.51	13.91	20.77	2.99	5.75	0.81	0.99	17.36	26.38	1.79	2.25	8.89	13.67	4.55	7.36
TCN$_{\text{seq}}$	16.88	28.67	3.77	6.83	2.03	4.84	0.06	0.24	14.85	23.42	3.60	5.06	4.30	4.90	0.66	0.99
GTS$_{\text{seq}}$	17.26	29.11	2.33	3.48	2.19	5.20	0.27	0.59	16.44	25.41	3.68	5.10	6.51	8.89	1.88	3.39
MTGNN$_{\text{seq}}$	19.88	32.94	7.83	12.68	2.12	4.75	0.38	0.44	14.86	22.58	2.59	3.61	10.26	14.92	1.16	1.81
PatchTST$_{\text{seq}}$	19.11	32.50	2.34	2.97	2.64	5.32	0.72	0.43	17.91	27.13	7.18	6.88	4.85	5.93	1.59	1.78
DLinear$_{\text{seq}}$	19.69	32.75	2.91	2.83	3.47	6.56	1.17	1.12	17.32	26.31	2.71	3.43	4.81	5.81	1.64	1.57
TimesNet$_{\text{seq}}$	17.77	29.91	3.13	6.93	3.92	7.18	1.46	2.51	18.38	27.61	4.33	5.15	5.18	6.13	1.72	2.03
iTransformer$_{\text{seq}}$	16.23	27.83	2.33	3.41	2.87	5.84	1.23	1.31	16.03	25.08	4.87	5.35	6.28	7.72	1.52	1.92
OFA$_{\text{seq}}$	19.10	32.48	2.21	2.43	3.04	6.33	1.26	1.57	17.40	26.20	5.32	3.69	4.72	5.22	1.63	1.85
KI-TSF$_{\text{seq}}$	17.30	29.38	4.38	7.80	2.02	4.73	0.30	0.50	14.73	23.31	3.91	5.07	4.70	5.85	1.97	3.46
KI-TSF	**15.23**	**25.32**	1.51	2.72	**1.75**	**4.46**	0.09	0.06	**13.41**	**21.30**	1.64	2.08	**3.24**	**4.24**	0.15	0.23

Representation-matching Sample Selection for Temporal Knowledge Coverage Maximization To tackle the forgetting of variable dependencies and temporal Knowledge in sequential training, we store a small subset of MTS samples and the structural knowledge from the previous regimes for memory replay when adapting the model to the current regime. Specifically, we propose an efficient sample selection scheme that maximizes the temporal and dependencies coverage of each regime given a limited memory budget. We first perform a distribution characterization by splitting the MTS data into the most diverse modes in the representation space, which can be formulated as a constrained optimization problem:

$$\begin{aligned} &\max_{0<K\le K_0}\ \max_{n_1,\cdots,n_K}\ \frac{1}{K}\sum_{1\le i\neq j\le K}\mathcal{D}\left(\mathcal{M}_i,\mathcal{M}_j\right) \\ &\text{s.t. } \forall i, \Delta_1 < |\mathcal{M}_i| < \Delta_2; \sum_i |\mathcal{M}_i| = n \end{aligned} \tag{1}$$

where $\mathcal{D}(\cdot,\cdot)$ can be any distribution-related distance metric, n is the number of training samples in a single regime, $\Delta1$, $\Delta2$ and K_0 are hyperparameters to avoid trivial partitions and over-splitting, $\mathcal{M}$ denotes the subset of representations that corresponds to contiguous samples. Specifically, we choose the Deep Correlation Alignment (CORAL) [13] to measure the temporal distribution similarity, *i.e.*, $D(\cdot,\cdot) = \frac{1}{4q^2}\left\|\mathrm{C}_{(\cdot)} - \mathrm{C}_{(\cdot)}\right\|_F^2$, where q is the number of dimensions for each hidden state, $\mathrm{C}(\cdot)$ denotes the second-order statistics (covariance matrix).

3 Experiment

3.1 Sequential Training and Testing Protocol

The experiments were conducted under the sequential training and testing protocol. At the training phase of i-th regime, the training objectives $\mathcal{L}_{\text{regime-i}}$ contains the objective for the current training samples, denoted as $\mathcal{L}_{\text{current}}$ and that for the memory of previous i-1 regimes, denoted as $\mathcal{L}_{\text{memory}}$. After training, the structural knowledge of the current regime is saved in the structural memory and the training samples are selected to enrich the MTS memory as aforementioned.

During the testing phase, KI-TSF successfully retains its ability to forecast for test sample queries from all previous and current regimes, and accordingly recover the learned dependency structures. We adopt two metrics based on Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) to evaluate the performance on continual MTS forecasting, *i.e.*, the Average Performance (AP) and Average Forgetting (AF) [9, 18].

3.2 Experiment Result

Table 1 summarizes the comparison results of selected baselines versus our proposed KI-TSF method for 12 horizon predictions on three public benchmark MTS datasets including the traffic (Traffic-CL), solar energy (Solar-CL), and human activity recognition (HAR-CL), as well as one synthetic dataset based on Non-repeating Random Walk [4] that is used in [8] under continual learning setting. Based on the experiment results in Table 1, we observe that statistical methods, such as VAR [10]and ARIMA [3], cannot perform well under continual learning settings and exhibit obvious performance degradation (AF) in sequential training (*i.e.*, seq).All deep learning based baseline methods, including state-of-art forecasting models, such as PatchTST [11], TimesNet [16], iTransformer [7], and even OFA [14] that equipped with the large language model (LLM), also suffer from obvious performance degradation (AF) in sequential training (*i.e.*, seq), suggesting the existence of catastrophic forgetting phenomenons when regime shifts. Finally, KI-TSF demonstrates the advantages over other baseline models, showing learning a dynamic structure is beneficial for MTS modeling, and the structural knowledge helps to characterize the general variable behaviors in each regime.

4 Conclusion

In this work, we propose a novel Knowledge Informed Time Series Forecasting (KI-TSF) framework to perform MTS forecasting and infer dependency structures in the continual learning setting. We develop a forecasting model based on dynamic graph learning and impose a consistency regularization that exploits structural knowledge to facilitate continual time series forecasting. We further exploit the temporal knowledge by proposing a novel representation-matching memory replay scheme, which maximizes the temporal coverage of MTS data to efficiently preserve each regime's underlying temporal dynamics and dependency structure. Experiments on one synthetic dataset and three real-world benchmark datasets demonstrate the effectiveness and advantages of the proposed KI-TSF on continual MTS forecasting tasks.

References

[1] Lei Bai, Lina Yao, Can Li, Xianzhi Wang, and Can Wang. 2020. Adaptive graph convolutional recurrent network for traffic forecasting. *Advances in Neural Information Processing Systems* 33 (2020), 17804–17815.

[2] Shaojie Bai, J. Zico Kolter, and Vladlen Koltun. 2018. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling. *arXiv:1803.01271* (2018).

[3] George EP Box, Gwilym M Jenkins, Gregory C Reinsel, and Greta M Ljung. 2015. *Time series analysis: forecasting and control*. John Wiley & Sons.

[4] Anne Denton. 2005. Kernel-density-based clustering of time series subsequences using a continuous random-walk noise model. In *Fifth IEEE International Conference on Data Mining (ICDM'05)*. IEEE, 8–pp.

[5] Aurora Gonzalez-Vidal, Fernando Jimenez, and Antonio F Gomez-Skarmeta. 2019. A methodology for energy multivariate time series forecasting in smart buildings based on feature selection. *Energy and Buildings* 196 (2019), 71–82.

[6] Bo Jin, Haoyu Yang, Leilei Sun, Chuanren Liu, Yue Qu, and Jianing Tong. 2018. A treatment engine by predicting next-period prescriptions. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 1608–1616.

[7] Yong Liu, Tengge Hu, Haoran Zhang, Haixu Wu, Shiyu Wang, Lintao Ma, and Mingsheng Long. 2023. iTransformer: Inverted Transformers Are Effective for Time Series Forecasting. *arXiv preprint arXiv:2310.06625* (2023).

[8] Yijing Liu, Qinxian Liu, Jian-Wei Zhang, Haozhe Feng, Zhongwei Wang, Zihan Zhou, and Wei Chen. 2022. Multivariate Time-Series Forecasting with Temporal Polynomial Graph Neural Networks. In *Advances in Neural Information Processing Systems*.

[9] David Lopez-Paz and Marc'Aurelio Ranzato. 2017. Gradient Episodic Memory for Continual Learning. In *NIPS*.

[10] Helmut Lütkepohl. 2005. *New introduction to multiple time series analysis*. Springer Science & Business Media.

[11] Yuqi Nie, Nam H Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. 2022. A time series is worth 64 words: Long-term forecasting with transformers. *arXiv preprint arXiv:2211.14730* (2022).

[12] Chao Shang and Jie Chen. 2021. Discrete Graph Structure Learning for Forecasting Multiple Time Series. In *Proceedings of International Conference on Learning Representations*.

[13] Baochen Sun and Kate Saenko. 2016. Deep CORAL: Correlation Alignment for Deep Domain Adaptation. In *ECCV Workshops*.

[14] Xue Wang Liang Sun Rong Jin Tian Zhou, Peisong Niu. 2023. One Fits All: Power General Time Series Analysis by Pretrained LM. In *NeurIPS*.

[15] Aäron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alexander Graves, Nal Kalchbrenner, Andrew Senior, and Koray Kavukcuoglu. 2016. WaveNet: A Generative Model for Raw Audio. (2016).

[16] Haixu Wu, Tengge Hu, Yong Liu, Hang Zhou, Jianmin Wang, and Mingsheng Long. 2022. Timesnet: Temporal 2d-variation modeling for general time series analysis. *arXiv preprint arXiv:2210.02186* (2022).

[17] Zonghan Wu, Shirui Pan, Guodong Long, Jing Jiang, Xiaojun Chang, and Chengqi Zhang. 2020. Connecting the dots: Multivariate time series forecasting with graph neural networks. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 753–763.

[18] Xikun Zhang, Dongjin Song, and Dacheng Tao. 2022. CGLB: Benchmark Tasks for Continual Graph Learning. In *Thirty-sixth Conference on Neural Information Processing Systems Datasets and Benchmarks Track*.