Learning Dynamic Temporal Relations with Continuous Graph for Multivariate Time Series Forecasting (Student Abstract)

Zhiyuan Wang[1], Fan Zhou[1,4], Goce Trajcevski[2], Kunpeng Zhang[3], Ting Zhong[1]
[1] University of Electronic Science and Technology of China, China
[2] Iowa State University, USA
[3] University of Maryland, College Park, USA
[4] Kashi Institute of Electronics and Information Industry, China
zhy.wangcs@gmail.com, gocet25@iastate.edu, fan.zhou@uestc.edu.cn, kpzhang@umd.edu, zhongting@uestc.edu.cn

Abstract

The recent advance in graph neural networks (GNNs) has inspired a few studies to leverage the dependencies of variables for time series prediction. Despite the promising results, existing GNN-based models cannot capture the global dynamic relations between variables owing to the inherent limitation of their graph learning module. Besides, multi-scale temporal information is usually ignored or simply concatenated in prior methods, resulting in inaccurate predictions. To overcome these limitations, we present CGMF, a Continuous Graph learning method for Multivariate time series Forecasting (CGMF). Our CGMF consists of a continuous graph module incorporating differential equations to capture the long-range intra- and inter-relations of the temporal embedding sequence. We also introduce a controlled differential equation-based fusion mechanism that efficiently exploits multi-scale representations to form continuous evolutional dynamics and learn rich relations and patterns shared across different scales. Comprehensive experiments demonstrate the effectiveness of our method for a variety of datasets.

Introduction

Multivariate time series (MTS) have important impacts in many aspects of daily life, and their forecasting plays a significant role on a plethora of modern applications, ranging from climate analysis, traffic and urban flows, to electricity industry and financial markets. A large body of research works have targeted the improvements of performance in MTS forecasting.

Recent advances in deep learning, especially graph neural network (GNN), spurred a few studies on modeling the MTS through exploiting the ability of GNN in learning neighborhoods' contexts. For example, recent graph-based models such as MTGNN (Wu et al. 2020) extract the unidirected or bidirectional relations among variables that captures the spatial-temporal dependencies within the time series.

Despite the encouraging results made by existing GNN-based works, there are certain issues that limits their performance: (1) Most GNN-based methods only consider the static graph (Wu et al. 2020) for memory efficiency – i.e., they employ the time-invariant graph structure between variables without the dynamic relations. (2) Time series in real world usually shows various patterns in multi-scale observations. Existing methods usually handle multi-scale representations by simple concatenation or linear units (Wu et al. 2020; Lin et al. 2019). In general, these operations can hardly learn informative cross-scale interactions.

In this paper, we propose a novel method, named **C**ontinuous **G**raph **M**ultivariate **F**orecasting (CGMF), aiming to address the aforementioned limitations. It establishes temporal dynamic correlations from multi-scale representations and exploits underlying connections between graphs and differential equations while inferring the temporal evolution. Thus, CGMF provides a direct way to simulate the dynamics of multivariate time series in the continuous space. Extensive experiments conducted on real-world data show that our model achieves the state-of-the-art results.

Methodology

Problem Definition. Given a MTS $\mathbf{X}_{1:T} = \{\mathbf{X}_t\}_{t\in[1:T]}$, where $\mathbf{X}_t = \{\mathbf{x}_t^{(i)}\}_{i\in[1:N]}$ denotes N variables at time t, and $\mathbf{x}_t^{(i)} \in \mathbb{R}^M$ is composed of M observations of i-th variable. Given $\mathbf{X}_{1:T}$, our goal is to produce the forecastings for each variable τ time steps ahead, given its historical observations, i.e., $\hat{\mathbf{Y}}_{T+\tau} = \mathrm{CGMF}(\mathbf{X}_{1:T})$.

Framework. There are three main modules in our proposed CGMF. The first module, employing the convolutional units, serves as an encoder that transform the MTS input into multi-scale embeddings for further processing. In the second component, we establish the dynamic temporal graph between variables and learn the long-range intra- and inter-relations over time series by introducing a continuous GNN. We treat the last component as the decoder, which fuses multi-scale representations to obtain the cross-scale knowledge and output forecastings.

Intra- and Inter-relations Learning. Unlike (Wu et al. 2020) that only considers static graph, we model time series with the dynamic temporal graph, which means the relationships between each variable is time-varying. Given multi-scale temporal embeddings $\{\mathbf{h}_1^{(i)}, \cdots, \mathbf{h}_{T_s}^{(i)}\} \in \mathbb{R}^{T_s\times N\times d}$ at scale s, we first reshape them into an 2-dimensional matrix as $\mathbb{R}^{T_s N\times d}$ and apply self-attention to obtain the score matrix $A \in \mathbb{R}^{T_s N\times T_s N}$. For memory efficiency, we keep the top k weights for each variable in matrix for further inference. Then, we incorporate ordinary differential equations

Method	Metric	Solor Energy 3	24	Traffic 3	24	Electricity 3	24
ARIMA	RSE	0.2546	0.8608	0.5998	0.6260	0.1008	0.1273
	CORR	0.9624	0.5387	0.7756	0.7463	0.8801	0.8590
GRU	RSE	0.2130	0.5063	0.5479	0.5804	0.1246	0.1420
	CORR	0.9753	0.8764	0.8360	0.8191	0.8665	0.8549
Informer	RSE	0.2391	0.7819	0.5476	0.5861	0.1479	0.1492
	CORR	0.9750	0.8895	0.8346	0.8057	0.8768	0.8692
MTGNN	RSE	0.1808	0.4459	0.4239	0.4672	0.0753	0.0955
	CORR	0.9802	0.9026	0.8893	0.8807	0.9447	0.9231
STODE	RSE	0.1952	0.4468	0.4785	0.5063	0.0879	0.1027
	CORR	0.9776	0.8859	0.8752	0.8587	0.9318	0.9126
CGMF	RSE	**0.1768**	**0.4310**	**0.4133**	**0.4526**	**0.0729**	**0.0931**
	CORR	**0.9835**	**0.9034**	**0.8942**	**0.8811**	**0.9478**	**0.9253**

Table 1: Performance comparisons across three real-world datasets. Best performance is in bold font.

into canonical GNN to aggregate node knowledge and the solver can be formulated as:

$$\text{ODE-Solver}(\mathbf{h}_t^0, l', L) = \mathbf{h}_t^0 + \int_0^L g_s(l', \mathbf{h}_t^{l'}; \bar{\mathbf{A}}) dl' \quad (1)$$

where $\mathbf{h}_t^0 = \mathbf{h}_t$ denoting the embedding at every scale, and $g_s(\cdot)$ can be any feed-forward network with the input $\bar{\mathbf{A}}\mathbf{h}_t^{l'}$. Here we let $g_s(\mathbf{h}_t^{l'}; \bar{\mathbf{A}}) = \bar{\mathbf{A}}\mathbf{h}_t^{l'}$ without additional learned parameters. It not only improves the efficiency in the computation of integral approximation but also aggregates the linear features among all nodes. After infinite aggregations, the representation $\mathbf{h}_t^L$ contains long-range correlations between nodes on the graph without over-smooth phenomenon.

Cross-scale Representation Fusion. We now pay attention to the fusion of representation at all scales. We propose a controlled differential equation-based method to better operate the cross-scale representation fusion. Given the multi-scale representations $\mathbf{H}$, we consider all representations and their intervals to simulate the evolutional process. The evolved representations $\mathbf{E} = \{\mathbf{e}_t\}_{t \in [1:T]}$ can be calculated by the following continuous dynamics:

$$\begin{aligned} \mathbf{e}_t &= \mathbf{z}_0 + \int_0^t g_c(t', \zeta(\mathbf{e}_{t'}, \{\mathbf{h}_{s,t'}\}_{s \in S})) d\mathcal{Z}_{t'} \\ &= \mathbf{z}_0 + \int_0^t g_c(t', \zeta(\mathbf{e}_{t'}, \{\mathbf{h}_{s,t'}\}_{s \in S})) \frac{d\mathcal{Z}_{t'}}{dt'} dt', \end{aligned} \quad (2)$$

where $\mathbf{z}_0$ is the initial state of the input time series $\mathbf{X}_{1:T}$ calculated by a backward GRU. $\mathcal{Z}'_{t'}$, consisting of a series of continuous vectors, are produced by the natural cubic spline (Kidger et al. 2020) with respect to $\mathbf{z}_{1:T}$ for the differentiability. With this equations, we can directly generate a continuous trajectory according to time and representations at all scale, which automatically extracts the trend, level and patterns of the time series across multiple scales. Finally, we use a MLP layer to make forecastings at future τ steps, i.e., $\hat{\mathbf{Y}}_{T+\tau} = \text{MLP}(\{\mathbf{H}, \mathbf{E}\})$.

Experiments

Datasets & Baselines. We conduct experiments on three public real-world datasets (Wu et al. 2020): Solar Energy, Traffic, and Electricity. We compare CGMF with the following baselines: ARIMA, GRU, Informer (Zhou et al. 2021), MTGNN (Wu et al. 2020), and STODE (Fang et al. 2021).

Performance Comparison. Table 1 summarizes the experimental results. We observe that CGMF achieves superiority and draw the following important conclusions: (1) Consistent with our expectations, long-term forecasting is more difficult since the model has to better understand levels, trends, and periodicity with respect to the time series. However, CGMF improves a lot and performs more stably in this situation. Our multi-scale embedding and fusion mechanisms extract the expressive cross-scale information that includes short- and long-term patterns, which effectively addresses this issue. (2) STODE also incorporates neural ODE into node relations learning module, but it is designed for spatial-temporal tasks (e.g., traffic flow prediction) and requires predefined structured information. Its non-linear ODE solver with learnable parameters is hard to converge, resulting in inaccuracy and time inefficiency.

Conclusion

In this paper, we proposed CGMF, a generative model that can disentangle three latent factors through constraining mutual information for better generalizing across non-stationary time series. Extensive empirical results demonstrated its superior performance over conventional domain generalization methods in both source and temporal generalization tasks. Additionally, the learned dynamic causal factors can improve the performance of conventional domain generalization methods under the non-stationary setting, because the dynamic and static factors can be mixed up easily and can be excluded from causal factors.

Acknowledgments

This work was supported in part by National Natural Science Foundation of China (Grant No.62072077 and No.62176043), Natural Science Foundation of Sichuan Province (Grant No. 2022NSFSC0505), and National Science Foundation SWIFT (Grant No.2030249).

References

Fang, Z.; Long, Q.; Song, G.; and Xie, K. 2021. Spatial-temporal graph ode networks for traffic flow forecasting. In *SIGKDD*, 364–373.

Kidger, P.; Morrill, J.; Foster, J.; and Lyons, T. 2020. Neural Controlled Differential Equations for Irregular Time Series. In *NeurIPS*, 6696–6707.

Lin, Z.; Feng, J.; Lu, Z.; Li, Y.; and Jin, D. 2019. Deepstn+: Context-aware spatial-temporal neural network for crowd flow prediction in metropolis. In *AAAI*, 1020–1027.

Wu, Z.; Pan, S.; Long, G.; Jiang, J.; Chang, X.; and Zhang, C. 2020. Connecting the dots: Multivariate time series forecasting with graph neural networks. In *SIGKDD*, 753–763.

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 *AAAI*, 11106–11115.