The Thirty-Seventh AAAI Conference on Artificial Intelligence (AAAI-23)

DyCVAE: Learning Dynamic Causal Factors for Non-stationary Series Domain Generalization (Student Abstract)

Weifeng Zhang[1], **Zhiyuan Wang**[1], **Kunpeng Zhang**[2], **Ting Zhong**[1], **Fan Zhou** [*] [1,3]

[1] University of Electronic Science and Technology of China, Chengdu, Sichuan 610054, China
[2] University of Maryland, College Park, MD 20742-0001, USA.
[3] Kashi Institute of Electronics and Information Industry, Kashi, Xinjiang 844000, China
weifzh@outlook.com, zhy.wangcs@gmail.com, kpzhang@umd.edu, zhongting@uestc.edu.cn, fan.zhou@uestc.edu.cn

Abstract

Learning domain-invariant representations is a major task of out-of-distribution generalization. To address this issue, recent efforts have taken into accounting causality, aiming at learning the causal factors with regard to tasks. However, extending existing generalization methods for adapting non-stationary time series may be ineffective, because they fail to model the underlying causal factors due to *temporal-domain shifts* except for source-domain shifts. To this end, we propose a novel model DyCVAE to learn dynamic causal factors. The results on synthetic and real datasets demonstrate the effectiveness of our proposed model for the task of generalization in time series domain.

Introduction

Out-of-distribution generalization task has been researched over decades, generally, machine learning paradigms often fail in test data which is not independent and identically distributed with training data (Torralba and Efros 2011). It is overreliance on correlation among features rather than causation that causes this issue. To address this problem, recently invariant causal representation studies (Arjovsky et al. 2019; Ahuja et al. 2021; Ilse et al. 2020) have gain more and more attention. Nevertheless, these models fail to work well for time series.

The phenomenon of temporal-domain shifts is ubiquitous, since the real world is constantly evolving. Additionally, the source-domain shifts still exist as discussed in domain generalization study. For instance, data may be collected from many heterogeneous sources, and different samples might be measured by different devices. Recent work (Gagnon-Audet et al. 2022) points out that the combination of these two distribution shifts leads conventional domain generalization methods ineffective. Therefore, it is necessary to construct a new method accounting for both two shifts in a unified viewpoint.

In this paper, we propose a novel model named *Dynamic Causal Factors Variational Auto-Encoder* (DyCVAE) based on variational inference, addressing the non-stationary time series generalization task. To be specific, we employ a deep generative model to learn three latent factors, including dynamic causal, dynamic non-causal, and static non-causal factors. Finally, extensive experiments with our model on synthetic and real data show that state-of-the-art results are achieved.

Methodology

Definition. Given the training data $\mathcal{D}^{train} = \{(\mathbf{x}^i_{1:T}, \mathbf{y}^i_{1:T})\}_{i=1}^K$, where $\mathbf{x}^i_{1:T} \in \mathcal{X}$ is the i-th non-stationary time series input, and $\mathbf{y}_i \in \mathcal{Y}$ is the corresponding labels, where T denotes the length of time series. We denote $P^{train}(\mathbf{x}, \mathbf{y})$ and $P^{test}(\mathbf{x}, \mathbf{y})$ as the distribution of training and test set, respectively. The goal is training a model $h : \mathcal{X} \to \mathcal{Y}$ whose risk is minimum on an unseen but related target domain $\mathcal{D}^{test}$: $R_{\mathcal{D}^{test}}(h) = \mathbb{E}_{(\mathbf{x},\mathbf{y})\sim P^{test}}[\ell(h(\mathbf{x}), \mathbf{y})]$, where $\ell : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}_+$ is a loss function.

Model architecture. We first separate latent factor into three latent factors $\mathbf{z}^c_{1:T}$, $\mathbf{z}^n_{1:T}$, $\mathbf{z}^s$, denoting dynamic causal, dynamic non-causal, and static non-causal factors, respectively. We assume these three factors are independent, thus the joint prior distribution can be factorized as

$$p_{\boldsymbol{\theta}}(\mathbf{z}) = p_{\boldsymbol{\theta}}(\mathbf{z}^s) \prod_{t=1}^{T} p_{\boldsymbol{\theta}}(\mathbf{z}^c_t|\mathbf{z}^c_{<t}) p_{\boldsymbol{\theta}}(\mathbf{z}^n_t|\mathbf{z}^n_{<t}). \tag{1}$$

The conditional priors of dynamic factors are defined as sequential priors and are implemented by two sequential neural network $p_{\boldsymbol{\theta}}(\mathbf{z}^c_t|\mathbf{z}^c_{<t}) = \mathcal{N}(\boldsymbol{\mu}(\mathbf{z}^c_{<t}), \boldsymbol{\sigma}(\mathbf{z}^c_{<t}))$ and $p_{\boldsymbol{\theta}}(\mathbf{z}^n_t|\mathbf{z}^n_{<t}) = \mathcal{N}(\boldsymbol{\mu}(\mathbf{z}^n_{<t}), \boldsymbol{\sigma}(\mathbf{z}^n_{<t}))$, and the prior of static factors is defined as a normal distribution $p_{\boldsymbol{\theta}}(\mathbf{z}^s) = \mathcal{N}(\mathbf{0}, \mathbf{I})$.

We then define a probabilistic generative model for the joint distribution over both observed and latent variables. Note that $\{\mathbf{x}_{1:T}, \mathbf{z}^s, \mathbf{z}^n_{1:T}\} \perp \mathbf{y}_{1:T}|\mathbf{z}^c_{1:T}$, we have following factorization:

$$p_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{y}, \mathbf{z}) = \underbrace{p_{\boldsymbol{\theta}}(\mathbf{y}_{1:T}|\mathbf{z}^c_{1:T})}_{\text{prediction}} \underbrace{p_{\boldsymbol{\theta}}(\mathbf{x}_{1:T}, \mathbf{z}^c_{1:T}, \mathbf{z}^n_{1:T}, \mathbf{z}^s)}_{\text{generation}}, \tag{2}$$

where the first term denotes the predictive part of dynamic causal factors $\mathbf{z}^c_{1:T}$. The causal factors learning was also adopted by recent domain generalization works (Lv et al. 2022). The second term denotes the generative process for non-stationary time series.

[*]Fan Zhou is the corresponding author.

Datasets	Fourier	TCMNIST	LSA64
ERM	9.55 ± 0.25	10.27 ± 0.14	48.78 ± 1.12
IRM	9.35 ± 0.09	10.04 ± 0.03	46.31 ± 1.51
IB-ERM	**10.08 ± 0.37**	9.99 ± 0.01	57.28 ± 1.88
IB-IRM	9.97 ± 0.61	10.05 ± 0.05	53.71 ± 1.95
VREx	9.74 ± 0.26	10.04 ± 0.03	46.11 ± 2.88
SD	9.70 ± 0.18	9.99 ± 0.00	50.74 ± 1.72
DIVA	9.60 ± 0.33	10.08 ± 0.31	58.24 ± 0.66
LSSAE	–	10.04 ± 0.05	–
DyCVAE	9.55 ± 0.41	**11.02 ± 0.83**	**61.97 ± 0.35**

Table 1: Overall performance comparisons.

Given observed data $\mathbf{x}$ and $\mathbf{y}$, our model employs variational inference strategy to learn an approximate posterior of three latent factors $q_{\boldsymbol{\phi}}$. We train it like how VAE (Kingma and Welling 2014) optimized. Generally, we maximize the likelihood and optimize the objective function of latent factors as:

$$\max_{\boldsymbol{\theta}} \mathbb{E}_{P^{tr}} \left[p_{\boldsymbol{\theta}}(\mathbf{x}_{1:T}, \mathbf{y}_{1:T}) \right]. \tag{3}$$

Inspired by the existing work about time series generation (Yingzhen and Mandt 2018; Zhu et al. 2020), there are two structures about the posterior distribution $q_{\boldsymbol{\phi}}$ over latent variables with regard to dynamic and static factors, i.e., full structure, which infers dynamic factors through static one; and factorized structure which infers dynamic factors without static factors. As for our proposed method, not only should we take dynamic and static relations into consideration but also separating causal and non-causal relations is needed. To achieve better disentanglement performance (Zhu et al. 2020), we adopt factorized structure to infer dynamic and static factors. The inference part of our proposed model is as follows:

$$q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x}) = q_{\boldsymbol{\phi}}(\mathbf{z}^s|\mathbf{x}_{1:T}) \prod_{t=1}^{T} q_{\boldsymbol{\phi}}(\mathbf{z}_t^c|\mathbf{x}_{<t}) q_{\boldsymbol{\phi}}(\mathbf{z}_t^n|\mathbf{x}_{<t}), \tag{4}$$

where the $q_{\boldsymbol{\phi}}(\mathbf{z}^s|\mathbf{x}_{1:T})$, $q_{\boldsymbol{\phi}}(\mathbf{z}_t^c|\mathbf{x}_{<t})$ and $q_{\boldsymbol{\phi}}(\mathbf{z}_t^n|\mathbf{x}_{<t})$ are Gaussian distributions parameterized by sequential models.

Experiments

Datasets & Baselines. To examine our proposed method, we conduct experiments on both synthetic and real-world data provided by a recent benchmark (Gagnon-Audet et al. 2022). We compare our proposed model with 8 baselines, which can be divided into two categories: model-agnostic methods, including ERM, IRM (Arjovsky et al. 2019), IB-ERM & IB-IRM (Ahuja et al. 2021), VREx (Krueger et al. 2021) and SD (Pezeshki et al. 2021); and deep generative methods including DIVA (Ilse et al. 2020) and LSSAE (Qin, Wang, and Li 2022). To ensure fair comparison, our model and all baselines share the feature extractor and predictor with the same hyper-parameters.

Performance Comparison. Since test domains are not available when training models in OOD setting, it is essential how to select the right model by validation. Here we adopt train-domain validation which is complied with domain generalization setting. The results of baselines and our proposed model are reported in Table 1. Note that LSSAE fails to work on datasets without label series. Overall, our proposed model outperforms all baselines on two datasets which are more complex than Fourier relatively. We attribute this improvement to learned dynamic causal factors.

Acknowledgments

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

References

Ahuja, K.; Caballero, E.; Zhang, D.; Gagnon-Audet, J.-C.; Bengio, Y.; Mitliagkas, I.; and Rish, I. 2021. Invariance principle meets information bottleneck for out-of-distribution generalization. In *NeurIPS*, volume 34.

Arjovsky, M.; Bottou, L.; Gulrajani, I.; and Lopez-Paz, D. 2019. Invariant risk minimization. *arXiv:1907.02893*.

Gagnon-Audet, J.-C.; Ahuja, K.; Darvishi-Bayazi, M.-J.; Dumas, G.; and Rish, I. 2022. WOODS: Benchmarks for Out-of-Distribution Generalization in Time Series Tasks. *arXiv:2203.09978*.

Ilse, M.; Tomczak, J. M.; Louizos, C.; and Welling, M. 2020. Diva: Domain invariant variational autoencoders. In *Medical Imaging with Deep Learning*, 322–348.

Kingma, D. P.; and Welling, M. 2014. Auto-Encoding Variational Bayes. In *ICLR*.

Krueger, D.; Caballero, E.; Jacobsen, J.-H.; Zhang, A.; Binas, J.; Zhang, D.; Le Priol, R.; and Courville, A. 2021. Out-of-distribution generalization via risk extrapolation (rex). In *ICML*, 5815–5826.

Lv, F.; Liang, J.; Li, S.; Zang, B.; Liu, C. H.; Wang, Z.; and Liu, D. 2022. Causality Inspired Representation Learning for Domain Generalization. *arXiv:2203.14237*.

Pezeshki, M.; Kaba, O.; Bengio, Y.; Courville, A. C.; Precup, D.; and Lajoie, G. 2021. Gradient starvation: A learning proclivity in neural networks. In *NeurIPS*, volume 34, 1256–1272.

Qin, T.; Wang, S.; and Li, H. 2022. Generalizing to Evolving Domains with Latent Structure-Aware Sequential Autoencoder. In Chaudhuri, K.; Jegelka, S.; Song, L.; Szepesvári, C.; Niu, G.; and Sabato, S., eds., *ICML*, volume 162, 18062–18082.

Torralba, A.; and Efros, A. A. 2011. Unbiased look at dataset bias. In *CVPR*, 1521–1528.

Yingzhen, L.; and Mandt, S. 2018. Disentangled sequential autoencoder. In *ICML*, 5670–5679.

Zhu, Y.; Min, M. R.; Kadav, A.; and Graf, H. P. 2020. S3vae: Self-supervised sequential vae for representation disentanglement and data generation. In *CVPR*, 6538–6547.