Rethinking the Parameter Learning of the Nonlinear Dynamical Probabilistic Latent Variable Model

Zhichao Chen, Hao Wang, Guofei Chen, Le Yao, Zhiqiang Ge, *Senior Member, IEEE* and Zhihuan Song

***Abstract*—Nonlinear dynamical probabilistic latent variable model (NDPLVM) and its variants have been widely applied in industrial data modeling scenarios like anomaly detection & diagnosis and inferential sensor. However, there remain two opening while long-neglected questions: (1) the parameter learning algorithm principle; (2) moment expressions for nonlinear neural network structure. To answer question 1), we propose a principled model named optimal control-NDPLVM (OC-NDPLVM) from stochastic differential equation theory and derive its parameter learning algorithm based on the alternating direction method of multipliers framework. To answer question 2), we conduct rigorous approximations for the mean and covariance equations. Based on the above-mentioned operations, we conduct two inferential sensor downstream tasks to demonstrate the effectiveness of the OC-NDPLVM.**

***Index Terms*—Probabilistic Latent Variable Model, Variational Inference, Convex Optimization, Machine Learning.**

***Note to Practitioners*—NDPLVMs have been widely applied in industrial process modeling to describe the nonlinearity, dynamics, and uncertainty of process measurement data. Despite various examples have proven the success of the NDPLVM, there remain two opening while long-neglected questions namely parameter learning principles and moment expressions to be answered. To alleviate these gaps, we propose the OC-NDPLVM model and its learning objective principally, derive its parameter learning algorithm based on the alternating direction method of multipliers framework, and propose the approximation of moment expressions in the model to make our computation more tractable. The above-mentioned operation answered the principle for NDPLVM parameter learning principles, inference network input, and moment expressions in the neural network. The optimization framework and approximation strategies proposed in this paper can be generalized to other DPLVM. And therefore, we suggest practitioners to adopt the methodology in this paper when they want to apply the NDPLVMs in industrial data modeling.**

I. Introduction

ACCURATE modeling the underlying dynamics of industrial data from high-dimensional sensory inputs is essential to intelligent manufacturing for the scenarios like decision-making and anomaly detection & diagnosis [1], [2]. To these ends, previous works adopt the dynamical probabilistic latent variable models (DPLVM) to model the nonlinearity, dynamics, and uncertainty of industrial process data. Throughout compressing the temporal pattern into the low-dimension markovian latent spaces, various downstream tasks like inferential sensor and anomaly detection & diagnosis can be constructed by harnessing these latent states.

Recently, with the development of representation learning, researchers tend to combine the DPLVM with deep learning architecture via designing neural network-based feature extraction structures. To simulate the nonlinear relationship between data and latent space, Fraccaro et. al propose [3] the Kalman Variational AutoEncoder model. To break up the Markov property assumption, recurrent neural networks (RNN) with memory cells like Long-Short Term Memory (LSTM) and Gated Recurrent Units (GRU) are therefore adopted in the network structure. On this basis, further works like variational recurrent neural network model [4] and Stochastic RNN model [5] are proposed. Meanwhile, to fit the multimodal distribution property of industrial data stemming from mode switching, Yao et. al [6] propose the variational inference framework under switching linear dynamical system framework. Quantities of works prove the success of the marriage between such DPLVM structure and neural networks.

Even though the feature extraction structures have been well discussed in previous works, the fundamental but essential model parameter learning part has not been widely discussed. Traditional parameter learning algorithm for the NDPLVMs is the amortized variational inference (AVI) algorithm. The AVI uses a neural network named inference network to infer the latent variable based on the observation data. On this basis, another neural network named generative network is designed to generate the observation data from the latent variable. After that, the model parameter learning is built upon the loss function consists of the regularization term (discrepancy between the inferred latent space and prior latent space), and the likelihood term (dicrepancy between the original data and generative data). When applying the AVI to the NDPLVMs, there remain the following vital but long-neglected questions:

1) **Parameter Learning Principles:** Since the data between different timestamps does not follow the i.i.d. assumption, and thus the factorization technique in the context of the Bayesian network is inevitably introduced to the data. This phenomenon results in a coupling learning objective, which hinders the parameter learning procedure design and the inference network input selection. How to design the parameter learning algorithm,

Manuscript received 2023. This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 92167106 and 62103362. (Corresponding author: Zhiqiang Ge, Zhihuan Song.)

Zhichao Chen, Hao Wang and Zhiqiang Ge are with the State Key Laboratory of Industrial Control Technology, College of Control Science and Engineering, Zhejiang University, Hangzhou 310027, China (e-mail: 12032042@zju.edu.cn; 22032130@zju.edu.cn; 3190105552@zju.edu.cn; gezq2023@163.com; songzhihuan@zju.edu.cn).

Le Yao is with the School of Mathematics, Hangzhou Normal University, Hangzhou 311121, China (email: yaole@hznu.edu.cn)