Rethinking the Parameter Learning of the Nonliear 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 sequential data modeling cases like anomaly detection & diagnosis and inferential sensor. However, previous works mainly concentrate on model architecture design, which may hinder their application for the ignorance of the following issues: (1), parameter learning algorithm principle; (2), inference network input variable; (3), moment expressions for nonlinear neural network structure. To address these issues, we propose a principled model named optimal control-NDPLVM (OC-NDPLVM), derive its parameter learning algorithm, and simplify its moment expressions in this paper. Specifically, we first propose the OC-NDPLVM and its parameter learning objective from stochastic differential theory principally. On this basis, we address issue 1) by converting the parameter learning problem into an optimization problem and deriving the parameter learning procedure based on the alternating direction method of multipliers framework. Meanwhile, we address issue 2) in the optimal control subproblem in the optimization procedure based on issue 1). After that, we address issue 3) by conducting the mean and variance expressions simplification to make the model computation tractable. Finally, we conduct two industrial inferential sensor downstream tasks to demonstrate the effectiveness of OC-NDPLVM.**

***Index Terms*—Probabilistic Latent Variable Model, Variational Inference, Stochastic Differential Equation, Sequential Data Modeling.**

***Note to Practitioners*—NDPLVMs have been widely applied in industrial process modeling to describe the nonlinearity, dynamic, and uncertainty. Despite various examples have proven the success of the NDPLVM, there remain challenges in parameter learning for applying the NDPLVM to industrial process modeling. 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 simplification of the model to make the 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 concerning simplification strategies proposed in this paper can be generalized to other DPLVM. And therefore, we suggest the practitioners to adopt the methodology in this manuscript when they are applying the NDPLVMs in industrail data modeling.**

Manuscript received 2022. This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 92167106 and 62103362. (Corresponding author: 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; gezhiqiang@zju.edu.cn; songzhihuan@zju.edu.cn).

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

I. INTRODUCTION

ACCURATE modeling the underlying dynamics of sequential 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 design the dynamical probabilistic latent variable models (DPLVM), to model the nonlinearity, dynamic, and uncertainty industrial process data. In the DPLVM, the temporal patterns like seasonality, trend, and residual in the sequential data are compressed into the low-dimension markovian latent states [3]. The 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 sophisticated approximate inference structures. To simulate the nonlinearity relationship between data and latent space, Fraccaro et. al propose [4] the Kalman Variational AutoEncoder model. To break up the Markov property assumption, recurrent neural networks (RNN) with memory cell like Long-Short Term Memory (LSTM) and Gated Recurrent Unit (GRU) are therefore adopted in the network structure. On this basis, further works like variational recurrent neural network (VRNN) model [5], Stochastic RNN (StoRNN) model [6] are proposed. Meanwhile, to fit the multimodal distribution property of industrial data stemming from mode switching, Yao et. al [7] 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 structure has 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. As a variant of variational inference algorithm, the AVI algorithm aims to find the maximum likelihood estimation of the model parameters for the observations sampled from a model consisting of latent variables (not observable) with unknown parameters. Specifically, 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 (difference between the