A DEEP NEURAL NETWORK APPROACH TO FAULT DETECTION IN STOCHASTIC NONLINEAR SYSTEMS

Kai Wang[a], Bhushan Gopaluni[b], Junghui Chen[c] and Zhihuan Song[d]

[a]The School of Automation, Central South University, Changsha, 410083, China
[b]Department of Chemical and Biological Engineering, The University of British Columbia,Vancouver, BC, Canada
[c]Department of Chemical Engineering, Chung Yuan Christian University, Chungli, Taoyuan, Taiwan, 32023, Republic of China
[d]State Key Laboratory of Industrial Control Technology, Zhejiang University, Hangzhou, 310027, Zhejiang, China

Abstract Overview

The dynamics of complex processes are characterized by strong nonlinearities and uncertainties. Monitoring such complex processes requires a high-quality model describing the corresponding nonlinear dynamic behavior. A deep neural network model is proposed to represent the state transition and observation equations in a standard stochastic nonlinear state space model. This model is learnt using an expectation-maximization algorithm and the posterior distributions of state variables are constructed by a forward-backward recurrent neural network. The resulting deep neural network model is used in detecting faults. The effectiveness of the proposed method is validated through the Tennessee Eastman (TE) process

Keywords

Nonlinear state space models, deep learning, dynamics, recurrent neural networks, fault detection

Introduction

Modern industrial plants are extremely complex due to high-dimensional variables, nonlinear dynamics and uncertainties. As such it is difficult to model and monitor these plants. Most existing process monitoring algorithms in the literature focus on one or two aspects of the complexities mentioned above. To account for process nonlinearity, different types of nonlinear mapping strategies have been proposed. Examples include Kernel PCA (Lee 2004) and manifold learning which are considered to be shallow methods due to their single-layer nonlinear mapping structure. Deep neural networks (DNNs) are state-of-the-art methods for approximating complex nonlinearities through multilayer nonlinear mapping. In recent years, different variants of DNNs such as stacked denoising auto-encoders (Zhang 2018), have been applied to industrial process monitoring.

A standard assumption with the above methods is that the data samples are identically and independently distributed, however, process data are serially correlated due to process dynamics and feedback control.

It is generally accepted that monitoring dynamic processes cannot be successfully conducted without accurately identifying dynamic models and their corresponding model accuracy. This observation is based on the experience of monitoring linear dynamic systems, where system identification tools like maximum likelihood estimation have been widely used. Using auto-encoding variational Bayes, also known as variational autoencoders (VAE) (Doersch 2016), in this abstract the nonlinear dynamic models are identified to perform nonlinear dynamic process monitoring. At the core of VAE is a nonlinear mapping from observations to latent variables (LVs) and the reverse mapping from LVs to observations using DNNs. The use of DNNs in VAE offers enough flexibility to approximate complex nonlinear mappings. An Expectation Maximization (EM) based parameter estimation algorithm is employed to train DNNs as this algorithm provides numerical stability and convergence. The VAE model is then extended to learn a nonlinear stochastic state-space model, which is approximated by a DNN.

Table 1. Process monitoring results of the TE process

	Fault detection rates[%]									
	DPCA		AR-DLV		DKPCA		CPM-DPCA		DNN-SS	
	T²	SPE	T²	SPE	T²	SPE	T²	SPE	T²	SPE
IDV1	99.4	99.9	93.1	99.9	99.5	99.8	99.5	99.5	80.1	99.9
IDV2	98.8	97.1	95.1	99.0	98.6	97.4	98.1	98.5	64.4	98.5
IDV3	9.7	4.8	8.3	7.1	9.4	4.6	2.4	3.3	**19.6**	14.4
IDV4	3.9	1.3	3.7	7.5	3.8	1.3	**13.8**	12.3	11.0	7.4
IDV5	27.1	8.3	24.2	27.5	27.0	17.4	35.7	38.5	31.4	32.8
IDV6	100	99.8	98.0	100	100	99.8	100	100	98.6	100
IDV7	45.4	21.2	42.5	37.4	45.6	25.9	49.7	53.6	52.5	51.4
IDV8	97.8	88.1	86.4	98.3	97.8	95.7	97.8	98.3	93.6	98.9
IDV9	1.8	1.4	2.1	1.8	1.9	1.4	1.3	2.4	2.5	2.4
IDV10	50.4	39.1	52.0	82.8	50.3	41.1	56.8	58.3	53.9	66.2
IDV11	19.4	53.1	7.0	47.5	18.5	x.4	23.8	29.7	18.6	47.5
IDV12	99.1	92.0	90.2	98.9	99.1	96.3	97.4	97.9	96.1	99.0
IDV13	94.2	94.5	93.5	95.5	94.2	94.9	94.5	94.7	94.9	94.5
IDV14	97.3	100	1.0	100	97.1	100	90.7	92.0	8.0	100
IDV15	12.8	1.9	6.7	9.8	9.6	1.9	29.8	28.2	**32.0**	16.6
	False alarm rates[%]									
IDV15	1.9	1.9	3.1	1.1	1.9	2.5	1.9	1.9	2.7	1.2

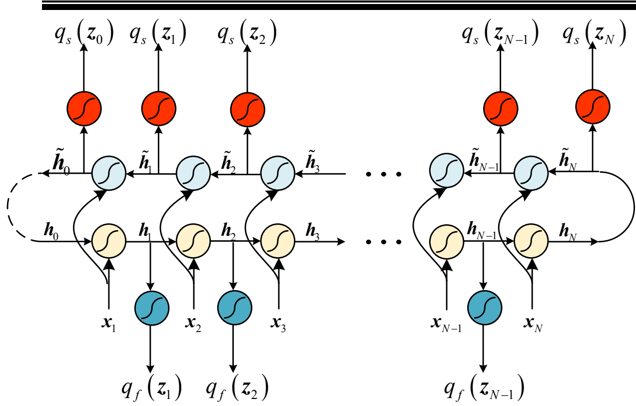

Fig. 1 The structure of forward-backward RNNs for learning the distribution of the posteriors.

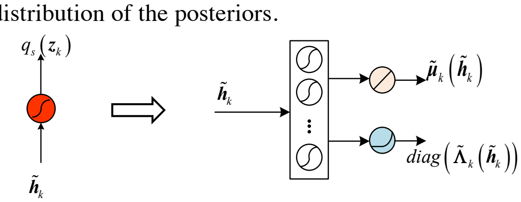

Fig. 2 The realization of the local Gaussian descriptor of smoothed distributions

Hence, the identified model is referred as a deep neural network within a stochastic nonlinear state space model, abbreviated as DNN-SS, and applied to process monitoring. After the model is learnt, monitoring indices can be readily designed based on the identified model as is common with the standard algorithms for process monitoring.

Learning DNN-SS Model

A stochastic discrete nonlinear LV model with a first-order Markov structure is considered as follows,

$$\begin{aligned} \boldsymbol{z}_k &= f\left(\boldsymbol{z}_{k-1}\right) + \boldsymbol{w}_k \\ \boldsymbol{x}_k &= g\left(\boldsymbol{z}_k\right) + \boldsymbol{v}_k \end{aligned} \qquad (1)$$

Note that $\boldsymbol{z}_k$ is also known as a state variable in dynamic systems. To avoid confusion, the phrase “state variables” is not used in this paper. Instead, $\boldsymbol{z}_k$ is referred as the LV while $\boldsymbol{h}_k$ in recurrent neural networks (RNN) specifies the cell state of RNN. In (1), $f\left(\boldsymbol{z}_{k-1}\right)$ is an unknown function that describes nonlinear dynamics. Given the corresponding LV, $g\left(\boldsymbol{z}_k\right)$ is another unknown nonlinear function that generates the observations (or measurements). $\boldsymbol{w}_k$ is unmeasured process noise which represents the process uncertainty while $\boldsymbol{v}_k$ represents the observation noise caused by sensors. Without any priors, $\boldsymbol{w}_k$ and $\boldsymbol{v}_k$ are assumed to be zero-mean Gaussian distributions, given by

$$p\left(\boldsymbol{w}_k\right) = N\left(0, I\right) \qquad (2)$$

$$p\left(\boldsymbol{v}_k\right) = N\left(0, \Gamma\right) \qquad (3)$$

where the covariance matrix of $\boldsymbol{w}_k$ can be assumed to be an identity matrix without any loss of generality. Simultaneously, the initial LV distribution is assumed to be

$$p\left(\boldsymbol{z}_0\right) = N\left(\boldsymbol{\mu}_0, \boldsymbol{\Lambda}_0\right) \qquad (4)$$

Thus, these terms $\boldsymbol{\theta}=\left\{\boldsymbol{\mu}_0, \boldsymbol{\Lambda}_0, f\left(\boldsymbol{z}_{k-1}\right), g\left(\boldsymbol{z}_k\right), \boldsymbol{\Gamma}\right\}$ are needed to be estimated for process modeling and monitoring applications. To learn the model using the expectation maximization (EM) algorithm, the lower bound on the likelihood function is initially rewritten as

$$\mathcal{L}(\boldsymbol{\theta}) = \int q(\boldsymbol{Z}) \ln p(\boldsymbol{X}, \boldsymbol{Z} \mid \boldsymbol{\theta}) d\boldsymbol{Z} - \int q(\boldsymbol{Z}) \ln q(\boldsymbol{Z}) d\boldsymbol{Z} \quad (5)$$

where $\boldsymbol{X} = \left\{\boldsymbol{x}_k \in R^m, k = 1, 2, \cdots, N\right\}$ is the training dataset and $\boldsymbol{Z} = \left\{\boldsymbol{z}_k \in R^n, k = 1, 2, \cdots, N\right\}$ is the corresponding unobserved LVs. $q\left(\boldsymbol{Z}\right)$ in (5) is chosen to be $p\left(\boldsymbol{Z} \mid \boldsymbol{X}, \hat{\boldsymbol{\theta}}\right)$ and is evaluated using the estimates from the previous EM iteration. The variational lower bound for optimizing the log likelihood is given by

$$\begin{aligned} &\arg\max_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}) = -kl\left(q(\boldsymbol{z}_0) \,||\, p(\boldsymbol{z}_0 \mid \boldsymbol{\theta})\right) \\ &-\sum_{k=1}^{N} E\left\{kl\left(q(\boldsymbol{z}_k \mid \boldsymbol{z}_{k-1}) \,||\, p(\boldsymbol{z}_k \mid \boldsymbol{z}_{k-1}, \boldsymbol{\theta})\right)\right\}_{\langle q(\boldsymbol{z}_{k-1})\rangle} \\ &+\sum_{k=1}^{N} E\left\{ln\, p(\boldsymbol{x}_k \mid \boldsymbol{z}_k, \boldsymbol{\theta})\right\}_{\langle q(\boldsymbol{z}_k)\rangle} + const \end{aligned} \quad (6)$$

where *const* stands for a constant term to be ignored. $E\{\bullet\}$ denotes the mathematical expectation operator and kl denotes the Kullback–Leibler (KL) divergence. To obtain the maximum likelihood estimate by maximizing (6), we need to define the posteriors $q(\boldsymbol{z}_k)$. Since the LV follows a first-order Markov model, the posteriors are obtained by a forward filter of the form $q_f(\boldsymbol{z}_k) = p(\boldsymbol{z}_k \mid \boldsymbol{x}_{1:k})$ and backward smoother of the form $q_s(\boldsymbol{z}_k) = p(\boldsymbol{z}_k \mid \boldsymbol{X})$ - similar to Kalman filter and smoother. The model nonlinearities and the corresponding posterior density functions are approximated using a RNN as shown in Fig. 1. The individual states of the cells in forward and backward RNN are used to approximate the distribution of the filtering and smoothing posteriors, respectively.

The distribution of LVs are approximated using the following local Gaussian descriptor applied at each time instance

$$q_s(\boldsymbol{z}_k) = p(\boldsymbol{z}_k \mid \boldsymbol{X}) = N\left(\tilde{\boldsymbol{\mu}}_k(\tilde{\boldsymbol{h}}_k), \tilde{\Lambda}_k(\tilde{\boldsymbol{h}}_k)\right) \quad (7)$$

DNN realization of the smoothed distribution (the red unit in Fig. 1) is presented in Fig. 2.

After the posterior distribution is implemented by the forward-backward RNNs, it is possible to learn the nonlinear dynamic system with the collected observation sequences $\boldsymbol{X} = \{\boldsymbol{x}_k, k = 1, 2, \cdots, N\}$. Hence, offline learning is performed with the given smoothed distribution $q_s(\boldsymbol{z}_k)$. In the offline learning stage, the first term in (6) will be $-kl\left(q_s(\boldsymbol{z}_0) \,||\, p(\boldsymbol{z}_0 \mid \boldsymbol{\theta})\right)$, implying the learned initial LV distribution $p(\boldsymbol{z}_0)$ should be close to the smoothed posterior $q_s(\boldsymbol{z}_0)$. This has been implemented by the backward RNN (Fig. 1), and $\tilde{\boldsymbol{h}}_0$ is the final output of the backward RNN. Also, $\tilde{\boldsymbol{h}}_0$ (the dashed line in Fig. 1) is assigned as the initial cell state of the forward RNN in each new iteration. However, the second and the third terms in the right side of (6) require the calculation of expectations associated with the posteriors. Because of complex distributions involved, it is difficult to evaluate the integrals in order to find the corresponding expectations. Therefore, an empirical average is used in place of the true expectation, i.e., the sampled value z_k^s drawn from the smoothed posterior $q_s(\boldsymbol{z}_k)$ is used to calculate the integrals. Based on VAE, DNNs can be used to represent $f(\boldsymbol{z}_{k-1})$ and $g(\boldsymbol{z}_k)$. Therefore, the model can be learned by mapping the observations into the distribution of LVs and reconstructing the observations from the sampled values from the distributions of LVs.

Case Studies

Five methods for online monitoring in the Tennessee Eastman (TE) process are compared, including dynamic PCA (DPCA)(Ku 1995), autoregressive-dynamic LV(AR-DLV)(Zhou 2017), dynamic KPCA(DKPCA)(Choi 2004) and constructive polynomial mapping DPCA(CPM-DPCA)(Yu 2017) and DNN-SS. In this example, the significance level α is 0.05. T^2 and SPE statistics are used to monitor process anomalies, respectively. Fault detection rate and false alarm rate are the two performance indices for process monitoring. Table 1 presents the comparative results for 15 faults in TE process. One can see the proposed modelling approach outperforms other approaches.

Conclusions

In this work, a novel DNN-SS is proposed to learn complex stochastic nonlinear state space model. The learning algorithm of DNN-SS is supported by VAE due to the general statistical efficiency based on variational Bayes and the general nonlinear representation allowed by DNNs. And the proposed modelling approach is applied to industrial fault detection. The experiments on the TE process validate the efficacy of the DNN-SS based monitoring framework.

References

Choi, S. W. and I.-B. Lee (2004). Nonlinear dynamic process monitoring based on dynamic kernel PCA. *Chemical Engineering Science* 59(24): 5897-5908.

Doersch, C. (2016). Tutorial on variational autoencoders. *arXiv preprint arXiv:1606.05908*.

Ku, W., R. H. Storer and C. Georgakis (1995). Disturbance detection and isolation by dynamic principal component analysis. *Chemometrics and intelligent laboratory systems* 30(1): 179-196.

Lee, J.-M., C. Yoo, S. W. Choi, P. A. Vanrolleghem and I.-B. Lee (2004). Nonlinear process monitoring using kernel principal component analysis. *Chemical Engineering Science* 59(1): 223-234.

Yu, H. and F. Khan (2017). Improved latent variable models for nonlinear and dynamic process monitoring. *Chemical Engineering Science* 168: 325-338.

Zhang, Z., T. Jiang, S. Li and Y. Yang (2018). Automated feature learning for nonlinear process monitoring – An approach using stacked denoising autoencoder and k-nearest neighbor rule. *Journal of Process Control* 64: 49-61.

Zhou, L., G. Li, Z. Song and S. J. Qin (2017). Autoregressive dynamic latent variable models for process monitoring. *IEEE Transactions on Control Systems Technology* 25(1): 366-373.