A Segmental Autoencoder-based Fault Detection for Nonlinear Dynamic Systems: An Interpretable Learning Framework

Hongtian Chen$^1$, Zhuofu Pan$^2$, Oguzhan Dogru$^1$, Yalin Wang$^2$, Biao Huang$^1$

1. University of Alberta, Edmonton T6G 2R3, Canada
2. Central South University, Changsha, 410083, China

E-mail: hongtian.chen@ualberta.ca; dogru@ualberta.ca; biao.huang@ualberta.ca

Abstract: This paper presents a segmental autoencoder-based fault detection (FD) framework for nonlinear dynamic systems. The basic idea behind the proposed FD scheme is to identify a generalized kernel representation based on the representation knowledge learned from an autoencoder. By using the system data, several cascades, linking nonlinear operators, are employed to obtain a data-based model which describes the nonlinear dynamic behaviors. With the help of the segmental structure of an autoencoder, a residual generator is then constructed. Rigorous mathematical analysis and an application on a continuous stirred tank reactor demonstrate the effectiveness of the proposed FD method.

Key Words: Data-driven Designs, Autoencoder, Fault Detection, Optimal Designs, Segmental Structures.

1 Introduction

Dynamic systems often involve complex interactions that are prompt to numerous uncertainties and abnormal behaviors [1]. When the nominal operation is affected by uncertainties or abnormal behaviors, system identification, monitoring and control become more challenging resulting loss of process safety, optimality, and economics [2–6]. Fault detection (FD) techniques provide systematic solutions to monitor the process and improve control performance [7–10].

Some common FD techniques utilize statistical information (e.g., correlation) between the input-output data by using linear models [5], such as principal component analysis [2, 11], support vector machine (SVM) [12–15]. Moreover, linear dynamic models have been developed through subspace identification [16–18]. Although these methods provide simple solutions to FD, their performance may be limited due to their linear structures [19].

Recent techniques utilize deep neural networks with nonlinear activation units to take the behavioral nonlinearity into consideration [20]. In the pattern recognition literature, these neural networks have also been used in anomaly/outlier/out-of-distribution detection [21], which are similar to the process monitoring and diagnosis techniques in terms of motivation and methodology. These data-driven approaches construct mappings between process input-output data to accurately estimate whether the operational conditions are within the desired ranges. Such an estimation is used to provide crucial information to the operator and improve the operational performance and safety.

In [20], a fully connected neural network and a recurrent neural network were employed to design the two FD methods, respectively corresponding to finite impulse response filters and a recursive filter. The study [15] proposed a variational autoencoder-based FD scheme by combining it with Kalman filter. More recently, both unsupervised and supervised learning tools were involved for system identification and the designs of residual generators in [22] from the view-point of interpretability. It is worth mentioning that, a perspective article [22] has proven the equivalence condition, based on which both unsupervised and supervised learning-based FD approaches can achieve the same FD performance.

Inspired by the study [22] where a bridge links unsupervised to supervised learning-based FD approaches, this work proposes a novel and optimal FD scheme for nonlinear dynamic systems. The proposed method consists of a segmental autoencoder network with stable kernel representations to detect the fault dynamically, without requiring prior process knowledge. Because it relies on nonlinear deep neural networks with a small number of trainable parameters, the computational efficiency is preserved during both the training and inference phases. The contributions of the study are summarized as follows:

- Formulating a data-based nonlinear model with the aid of the cascades to capture system dynamics,
- Constructing a generalized kernel representation using a segmental autoencoder,
- Developing a novel FD algorithm to detect the faults more accurately,
- Rigorously deriving analytical expressions to prove optimality of the proposed FD method,
- Quantitatively analyzing the performance of the proposed scheme through a continuous stirred tank reactor (CSTR) simulation.

The rest of this article is organized as follows. Detailed background information and problem formulation are provided in Section 2. The proposed FD method using the segmental autoencoder and performance analysis are detailed in Section 3. Simulation studies are presented in Section 4, and the conclusion is given in Section 5.

2 Preliminaries and Problem Formulation

2.1 Autoencoder

An autoencoder is a typical kind of unsupervised learning methods, which learns effective codings from unlabeled data. Thanks to its ability to capture important information and features, autoencoders have been intensively developed over the past four decades, and their variants include variational,
denoising, sparse, concrete autoencoders, etc. [23].

![Fig. 1: Schematic of an autoencoder](image)

As shown in Fig. 1, an autoencoder consists of two nonlinear mappings that correspond to the encoder and the decoder. Mathematically, it can be described by

\[ \ell = \phi_{\text{en}} h, \quad \hat{h} = \phi_{\text{de}} \ell, \quad \text{or} \quad \hat{h} = (\phi_{\text{de}} \circ \phi_{\text{en}}) h, \]

and its objective function \( L_{\text{total}} \) is

\[ L_{\text{total}} = \| h - \hat{h} \|_2^2, \]

where \( \phi_{\text{en}} \) and \( \phi_{\text{de}} \) are the mappings generated by the encoder and decoder, respectively; \( h \) and its estimation \( \hat{h} \) are the input and output signals, and \( \ell \) is the latent variable of a reduced dimension; “\( \circ \)” represents the cascade connection between two nonlinear operators.

### 2.2 Problem Description

Consider a nonlinear system that is defined as

\[
\begin{align*}
x(k+1) &= f(x(k), u(k), w(k)), \\
y(k) &= g(x(k), u(k), v(k)),
\end{align*}
\]

where \( x \in \mathbb{R}^n \), \( u \in \mathbb{R}^l \), and \( y \in \mathbb{R}^m \) are the system state, input, and output, respectively; \( w \in \mathbb{R}^p \) and \( v \in \mathbb{R}^q \) are noise sequences; \( f(\cdot) \) and \( g(\cdot) \) are continuous nonlinear mappings. Since only \( u \) and \( y \) of (3) can be used, a moving horizon (i.e., the stacked vector [22]) is usually adopted for system identification:

\[
\begin{align*}
u_p(k) &= [u^T(k-s_p) \cdots u^T(k-1)]^T, \\
u_f(k) &= [u_f^T(k) \cdots u_f^T(k+s_f)]^T.
\end{align*}
\]

The notation above can be used for any variable in (3).

In this paper, we intend to design a data-driven FD scheme for dynamic systems given in (3) with the help of an autoencoder. Considering the actuator fault \( \theta_a \) and sensor fault \( \theta_s \), (3) becomes [3]

\[
\begin{align*}
x(k+1) &= f(x(k), u(k), w(k), \theta_a(k)), \\
y(k) &= g(x(k), u(k), v(k), \theta_s(k)) + \theta_a(k).
\end{align*}
\]

For the FD purpose, the generalized version of a stable kernel representation of (3) is introduced below [20].

**Definition 1.** Given a nonlinear system (3), a nonlinear operator \( K \) is called a generalized stable kernel representation of (3) if for \( w = 0, v = 0 \), and a given \( x(0) \), the following holds

\[
\theta_a = 0, \theta_s = 0 \implies \left[ \begin{array}{c}
y_{\text{mix}}(k) \\
g_{\text{f}}(k)
\end{array} \right] = 0,
\]

where “\( [\cdot] \)” represents the composite operator, and \( z_{\text{mix}}(k) \) is the mixture of the stacked \( u \) and \( y \):

\[
z_{\text{mix}}(k) = [u_p^T(k) \ u_f^T(k) \ y_p^T(k)]^T.
\]

Based on the nonlinear \( K \) defined in (6), the residual signal can be obtained according to

\[
r(k) = y_{\text{f}}(k) - \hat{y}_{\text{f}}(k),
\]

where the estimation \( \hat{y}_{\text{f}}(k) \) is

\[
\hat{y}_{\text{f}}(k) = P z_{\text{mix}}(k).
\]

Based on the generalized stable kernel representation given in Definition 1 and the residual signal obtained via (8), two objectives of this study are summarized as follows:

- To obtain an analytical solution to the data-based model of (1) by using the cascade and composite operators;
- To propose a novel data-driven nonlinear FD strategy with the help of a segmental autoencoder and prove its optimal FD power.

### 3 Optimal Fault Detection Using An Segmental Autoencoder

#### 3.1 Data-based Model

In order to establish a data model of (3), we divide both \( f(\cdot) \) and \( g(\cdot) \) into three segments, i.e.,

\[
\begin{align*}
f_x : (k) \rightarrow x(k+1), \\
f_u : u(k) \rightarrow x(k+1), \\
f_w : w(k) \rightarrow x(k+1), \\
g_x : x(k) \rightarrow y(k), \\
g_u : u(k) \rightarrow y(k), \\
g_v : v(k) \rightarrow y(k),
\end{align*}
\]

where the subscripts correspond to the inputs of the mappings. While a general case will be considered shortly, in the special case, each variable on the right hand side of (3) has additive contributions to the left hand side. Using the future horizon \( s_f \) and combining (1) with (10) yields

\[
g_{s_f(k)} = P_x x(k) + P_u u_{s_f}(k) + P_w w_{s_f}(k) + v_{s_f}(k),
\]

where \( P_x, P_u, \) and \( P_w \) are the composite operators, and their forms are given as follows.

\[
P_x = \begin{bmatrix} g_x \\ g_x \circ f_x^1 \\ \vdots \\ g_x \circ f_x^{s_f} \end{bmatrix} : \mathbb{R}^n \rightarrow \mathbb{R}^{(s_f+1)n},
\]

\[
P_u = \begin{bmatrix} g_u \\ g_u \circ f_u^{s_f-1} \\ \vdots \\ g_u \circ f_u \end{bmatrix} : \mathbb{R}^{(s_f+1)l} \rightarrow \mathbb{R}^{(s_f+1)m},
\]

\[
P_v = \begin{bmatrix} g_v \\ g_v \circ f_v^{s_f-1} \\ \vdots \\ g_v \circ f_v \end{bmatrix} : \mathbb{R}^{(s_f+1)q} \rightarrow \mathbb{R}^{(s_f+1)q},
\]

\[
P_x | P_u | P_w : \mathbb{R}^{(s_f+1)n} + \mathbb{R}^{(s_f+1)m} + \mathbb{R}^{(s_f+1)q} \rightarrow \mathbb{R}^{(s_f+1)n} + \mathbb{R}^{(s_f+1)m} + \mathbb{R}^{(s_f+1)q}.
\]
In (12) to (14), $f^{sf}_x$ is defined by $f^{sf}_{x-2} \circ f_x \circ f_x = f^{sf}_{x-1} \circ f_x = f^{sf}_x$.

### 3.2 Data-based State Estimation

Since the system state, $x$, is unknown, the data-based model (11) is tractable. As an estimation, $\hat{y}$ can be obtained through a nonlinear estimator $\alpha$, such as a mapping generated by neural networks. By using the past horizon $s_p$, one can obtain

$$e(k) := y(k) - \hat{y}(k) = y(k) - g_x \circ \hat{x}(k) + g_u u(k), \quad \hat{x}(k+1) = f_x \circ \hat{x}(k) + f_u u(k) - o \circ e(k),$$

where $o : R^m \rightarrow R^n$. It can be derived as follows

$$\hat{x}(k) = f^{sr}_x \hat{x}(k-s_p) + O_{total} \hat{z}_p(k), \quad z_p(k) = [u_p(k) y_p(k)]^T, \quad O_{total} = [O_u O_y],$$

in which $O_u$ and $O_y$ have the similar structures as $P_u$ and $P_y$. Since $o$ is a nonlinear estimator such that $f^{sr}_x \hat{x}(k-s_p) \approx 0$ for a large $s_p$, the state estimation can be achievable, i.e.,

$$\hat{x}(k) \approx O_{total} \hat{z}_p(k).$$

Substituting (17) into (11) yields

$$y_{sf}(k) - P_y y_{sf}(k) - P_x \circ O_{total} \hat{z}_p(k) - P_u u_{sf}(k) = 0,$$

where $P_x$, $P_u$, and $P_y$ are

$$P_x = \begin{bmatrix} g_x & \vdots & \vdots & \vdots & 0 \\
\vdots & 0 & \vdots & \vdots & \vdots \\
g_x \circ f^{x}_{x+1} & \vdots & \vdots & \vdots & 0 \end{bmatrix}, \quad P_y = \begin{bmatrix} g_x & \vdots & \vdots & \vdots & 0 \\
\vdots & 0 & \vdots & \vdots & \vdots \\
g_x \circ f^{x}_{x+1} \circ o & \vdots & \vdots & \vdots & 0 \end{bmatrix}, \quad P_u = \begin{bmatrix} g_u & \vdots & \vdots & \vdots & \vdots \\
\vdots & 0 & \vdots & \vdots & \vdots \\
g_x \circ f^{x}_{x+1} \circ f_u & \vdots & \vdots & \vdots & g_x \end{bmatrix}$$

In (19), $f^{sf}_x = f_x - o \circ g_x$ and $f^{sf}_u = f_u - o \circ g_u$.

### 3.3 A Segmental Autoencoder-based $K$

Based on (18), the objective function of identification of $K$ and the design of residual generator are formulated as

$$\min || e_{sf} ||_2 \Rightarrow K^* := \arg \min_K || e_{sf} ||^2,$$

in which

$$e_{sf} = \left[ P_x \circ O_{total} P_y \right] - I \begin{bmatrix} z_{mix}(k) \\ y_f(k) \end{bmatrix} = -y_f(k) + \left[ P_x \circ O_{total} P_u P_y \right] \begin{bmatrix} z_{mix}(k) \\ y_f(k) \end{bmatrix}.$$

For nonlinear system (3), $K$ is replaced by a full connected neural network, i.e., a segmental autoencoder in this study. Based on the objective given in (20), the optimal hyperparameter of a segmental autoencoder can be obtained, equivalent to $K^*$.

Motivated by (20), the design concept, together with optimal FD performance, of the proposed FD scheme is summarized in the following theorem.

**Theorem 1.** Given the nonlinear system (3), the residual generator is constructed by

$$r(k) = y_{sf}(k) - \mathcal{H}_y(T^*) \begin{bmatrix} z_{mix}(k) \\ y_f(k) \end{bmatrix},$$

where $\mathcal{H}$ and $T$ are the structure and hyper parameter of an autoencoder, respectively. $\mathcal{H}(T^*)$ and its segment $\mathcal{H}_y(T^*)$ are obtained by

$$\arg \min_{\mathcal{H}(T^*)} || e_{total} ||^2_2,$$

based on which $\mathcal{H}_y(T^*)$ is

$$e_{total} = \begin{bmatrix} e_{mix} \\ e_{sf} \end{bmatrix}, \quad \begin{bmatrix} \hat{z}_{mix}(k) \\ y_f(k) \end{bmatrix} = \mathcal{H}_y(T^*) \begin{bmatrix} \mathcal{H}_y(\mathcal{Y}) \\ y_f(k) \end{bmatrix}.$$

Then $T^2$ defined on $r(k)$

$$T^2(r(k)) = r^T(k) \Sigma^{-1} r(k), \quad \Sigma = \mathcal{E}(rr^T),$$

has optimal FD power, where $\mathcal{E}(\cdot)$ is the expectation operator.

**Proof.** Based on the autoencoder $\mathcal{H}(T)$ defined in (24), (23) indicates

$$\arg \min_{\mathcal{H}(T^*)} || e_{total} ||^2_2 \Rightarrow \arg \min_{\mathcal{H}_y(T^*)} || e_{sf} ||^2_2.$$

With the parameter (or called weight) $\mathcal{H}_y(T^*)$, the residual generator given in (22) results in

$$\min || e_{sf} ||^2_2 = \min || r ||^2_2 \Rightarrow \min Tr(\Sigma),$$

where $Tr(\cdot)$ is the trace operator. Combining (5), (22) in the faulty condition becomes

$$r^T(k) = r(k) + f_{rel},$$

where $f_{rel}$ includes $\theta_u$ and $\theta_s$-related terms, whose mathematical description is given in [22]. Correspondingly, (25) will be

$$T^2(r^T(k)) = T^2(r(k)) + f_{rel}^T \Sigma^{-1} f_{rel},$$

because, in general, $\theta_u$ and $\theta_s$ are independent of $e_{sf}$. By combining (27) with (29), it can be verified that the segment $\mathcal{H}_y(T^*)$ and the associated $K^*$ can maximize the influence of $f_{rel}^T \Sigma^{-1} f_{rel}$, indicating the optimal FD power of the proposed scheme.

Therefore, this theorem is proven.
It is worth mentioning that when neural networks are employed to design an FD scheme, the fault-related term $f_{rel}$ must not disappear or be weakened. In addition, $f_{rel}$ can be quantitatively described. For example, the study [24] considered this problem based on the following Taylor series expansion:

$$r^f \approx r + \nabla r \Delta \theta - \frac{1}{2} (\Delta \theta)^T H \Delta \theta,$$

where $\Delta \theta$ is the changes caused by $\theta_x$ and $\theta_y$. Since $r$ is obtained via the optimal hyperparameter $H(Y^*)$, one can obtain the following two results:

$$\nabla r = 0 \text{ and } H \neq 0,$$

where $H$ is the Hessian matrix of $H(Y)$. Therefore, we have

$$f_{rel} = -\frac{1}{2} (\Delta \theta)^T H \Delta \theta,$$

and

$$f_{rel}^T \Sigma^{-1} f_{rel} > 0 \implies \mathcal{E}(T^2(r^f(k))) > \mathcal{E}(T^2(r(k))).$$

Here, $H$ is a three-dimensional tensor, ensuring that the fault influences will exist in the obtained residual signal.

![Fig. 2: The proposed segmental autoencoder-aided FD architecture](image)

The architecture of the proposed segmental autoencoder-aided FD scheme is depicted in Fig. 2. We can observe that only the segment of autoencoder is used to construct residual signals. It is the main difference between the proposed scheme and the existing autoencoder-based FD approaches.

### 3.4 Performance Comparison and Implementation

In order to provide a more detailed analysis, performance comparisons with the traditional autoencoder-based FD strategy and some possible variants will be conducted. For this purpose, we choose a semi-autoencoder (SmAE), a total autoencoder (ToAE), latent variables of a total-autoencoder (LaToAE)-based approaches as examples.

Let us consider SmAE and ToAE first. Rewriting (24) can obtain the following equations:

$$e_{\text{total}} = \begin{bmatrix} e_{\text{mix}} \\ e_{\text{sf}} \\ e_{\text{semi}} \end{bmatrix} = \begin{bmatrix} e_p \\ e_{sf,u} \\ e_{sf,v} \end{bmatrix} = \begin{bmatrix} e_p \\ e_{sf,u} \\ e_{sf,v} \end{bmatrix}.$$  

Therefore, two covariance matrices, corresponding to SmAE and ToAE-based approaches, can be obtained via

$$\Sigma_{\text{semi}} = \mathcal{E}(r_\text{semi}^T r_\text{semi}), \Sigma_{\text{total}} = \mathcal{E}(r_\text{total}^T r_\text{total}),$$

where $r_\text{semi}$ and $r_\text{total}$ are defined by

$$r_\text{semi} = r_\text{semi}^T, r_\text{total} = r_\text{total}^T.$$

By using SmAE with the structure $H_\text{semi}(Y^*)$ and ToAE with the structure $H(Y^*)$, two residual generators given in (36) are constructed by

$$\begin{bmatrix} z_{\text{mix}}(k) \\ y_f(k) \end{bmatrix} = \begin{bmatrix} H_{\text{sf}}(Y) \\ H_{\text{semi}}(Y^*) \end{bmatrix} \begin{bmatrix} z_{\text{mix}}(k) \\ y_f(k) \end{bmatrix}$$

where $H_{\text{semi}}(Y^*)$ and $H(Y^*)$ are obtained according to the same objective function as $H(Y^*)$, i.e.,

$$\arg min_{H} \mathcal{L}_\text{total} = \arg min_{H} \| e_{\text{total}} \|_2^2,$$

$$\implies \arg min_{H} \mathcal{L}_\text{total} \|_2^2 \implies \arg min_{H} \mathcal{L}_\text{total} \|_2^2.$$

Because of

$$\| e_{\text{total}} \|_2^2 > \| e_{\text{semi}} \|_2^2 > \| e_{\text{sf}} \|_2^2,$$

the following relationship holds

$$\text{Tr}(\Sigma_{\text{total}}) > \text{Tr}(\Sigma_{\text{semi}}) > \text{Tr}(\Sigma),$$

where $\Sigma_{\text{total}}$ and $\Sigma_{\text{semi}}$ are the covariance matrices of $r_\text{total}$ and $r_\text{semi}$, respectively. Combining (29) with (40) reveals the proposed scheme can achieve the best FD performance.

Let $r_{\text{latent}}$ be a feature indicator generated by LaToAE. It is easy to know that

$$\| r_{\text{latent}} \|_2^2 \approx \left\| \begin{bmatrix} z_{\text{mix}}(k) \\ y_f(k) \end{bmatrix} \right\|_2^2 \gg \| e_{\text{total}} \|_2^2,$$

based on which, we have

$$\text{Tr}(\Sigma_{\text{latent}}) \gg \text{Tr}(\Sigma_{\text{total}}),$$

where $\Sigma_{\text{latent}}$ is the covariance matrix of $r_{\text{latent}}$. It directly results in the conclusion that when the latent variables of autoencoder are used as feature indicators, faults are more difficult to detect because the latent variables carry the most of data variation information (i.e., the largest uncertainties).

Therefore, the aforementioned mathematical analysis has illustrated the optimal FD power of the proposed segmental autoencoder-based scheme.

According to Fig. 2 and Theorem 1, the complete procedures of the segmental autoencoder used for FD are summarized in Algorithms 1 and 2.
Algorithm 1 The segmental autoencoder-based FD application: Off-line learning

1: By using multiple unit delays, both the input and reference signals of an autoencoder are formed according to (24);
2: Construction of an autoencoder with the structure $H(T)$ is followed;
3: Based on (22), the residual signal $r(k)$ can be formulated with the help of the segment $H_{r}(T^*)$;
4: $T^2$ test statistic is obtained based on (25);
5: The threshold of $T^2$ can be determined by $J_{th} = \chi^2_{\alpha}(\sigma + 1)m$, where $\beta$ is the significance level.

Algorithm 2 The segmental autoencoder-based FD application: Online test

1: Read the online input $u_{(online)}$ and the output $y_{(online)}$;
2: Compute $r_{(online)}$ according to (22);
3: Obtain the online test statistic based on (29);
4: Perform an FD decision as follows:
   
   Fault alarm, if $T^2(r_{(online)}) > J_{th}$; \hspace{1cm} (43)
   
   Go back to step 1, otherwise.

4 Case Studies

4.1 Continuous Stirred Tank Reactor

A CSTR can be modeled by using a set of dynamic differential equations [25]. Fig. 3 shows the sketch of the facility, where $T$, $Q$, and $C$ represent the reactor temperature, flow rate, and concentration, respectively.

The system input $u$ and output $y$ can be defined by

$$u = [T_0, T, C]^T, \quad y = [T_0, T, C, Q_i]^T. \hspace{1cm} (44)$$

A total of $1.32 \times 10^4$ data samples are obtained from the CSTR simulation. These data are divided into a training data set with $1.2 \times 10^3$ samples and an online test data set with $1.2 \times 10^3$ samples. For the construction of a segmental autoencoder, 75 unit-delays are introduced to the input and 75 unit-delays are introduced to the reference output; the hidden layer has 95 neurons. The main configurations of the autoencoder used in this study are summarized in Table 1.

| No. of inputs | 77 |
|-------|-----|
| No. of outputs | 77 |
| No. of neurons in hidden layers | 38, 19, 38 |
| Activation functions | Gaussian, Affine, Gaussian, Affine |
| Optimizer | RMSprop |
| Learning rate | $1 \times 10^{-4}$ |
| Epoch | 24 |
| Batch | 16 |

Table 1: Configurations of the autoencoder used in constructing residual generators

4.2 FD Results Using The Segmental Autoencoder

Three faults occurring from the 201-st step, denoted as $\theta_a$, $\theta_{a1}$, and $\theta_{a2}$, are adopted in this case study. Following the description given in [25], $\theta_a$ is a structure (and multiplicative) fault in $\omega_b$, which is set to $\theta_a = \omega_b \cdot e^{-0.001(k-200)}$; $\theta_{a1}$ is an additive fault in $C$, whose value is $\theta_{a1} = 0.001(k-200)$; $\theta_{a2}$ is an additive fault in $T$ of the amplitude $\theta_{a2} = 0.05(k-200)$.

The FD performance for normal and faulty cases is evaluated respectively via false alarm ratio (FARs) and missing detection rates (MDRs). The two evaluation indices are calculated or estimated through

$$\text{FAR} = \text{Pr}(T^2(r) > J_{th}|\theta = 0),$$
$$\text{MDR} = \text{Pr}(T^2(r) \leq J_{th}|\theta \neq 0), \hspace{1cm} (45)$$

where $\theta = \{\theta_a, \theta_{a1}\}$, and $\text{Pr}()$ represents the probability function.

The FD results for three kinds of faults are shown in Fig. 4, where the blue solid lines are $T^2$ test statistics and the red dashed ones are the threshold $J_{th} = 48.35$. It can be
observed from Fig. 4 that the drifted faults contribute to a gradual increase of test statistics. Specifically, FARs are 0%, 0.52%, and 0.52%, and MDRs are 9.49%, 0%, and 2.22%; they respectively correspond to the results in Fig. 4 from the top to the bottom. Overall, the proposed segmental autoencoder-based approach shows satisfactory results in the simulation of various FD tasks.

5 Conclusion

This paper has proposed an optimal FD scheme for nonlinear systems, whose core is an unsupervised learning strategy that can identify nonlinear system dynamics. By adopting a segmental autoencoder, the minimized reconstruction error can be expressed by explainable knowledge representation. Furthermore, multiple unit delays are equipped in the autoencoder to improve the computation efficiency and the stability of learning procedures. Then, a residual generator can be constructed for the FD purpose. The validation studies have shown that the proposed FD method effectively detects faults for nonlinear dynamic systems.

References

[1] D. Cheng, R. Ortega, and E. Panteley, Advanced Robust and Adaptive Control—Theory and Applications. Beijing: Tsinghua University Press, 2005, chapter 1.

[2] B. Huang and R. Kadali, Dynamic Modeling, Predictive Control and Performance Monitoring: A Data-driven Subspace Approach. London: Springer-Verlag, 2008, chapter 3.

[3] S. X. Ding, Advanced Methods for Fault Diagnosis and Fault-tolerant Control. Berlin: Springer Nature, 2020, chapter 10.

[4] S. J. Qin, An overview of subspace identification, Computers & Chemical Engineering, 30(10-12): 1502-1513, 2006.

[5] H. Chen, B. Jiang, S. X. Ding, and B. Huang, Data-driven fault diagnosis for traction systems in high-speed trains: A survey, challenges, and perspectives, IEEE Transactions on Intelligent Transportation Systems, 2020, doi:10.1109/TITS.2020.3029946.

[6] Y. Song, Z. Liu, A. Rimnquist, P. Nāvik, and Z. Liu, Contact wire irregularity stochasticities and effect on high-speed railway pantograph-catenary interactions, IEEE Transactions on Instrumentation and Measurement, 69(10): 8196-8206, 2020.

[7] R. Chi, H. Zhang, B. Huang, and Z. Hou, Quantitative data-driven adaptive iterative learning control: From trajectory tracking to point-to-point tracking, IEEE Transactions on Cybernetics, 2020, doi:10.1109/TCYB.2020.3015233.

[8] D. Zhou, Y. Zhao, Z. Wang, X. He, and M. Gao, Review on diagnosis techniques for intermittent faults in dynamic systems, IEEE Transactions on Industrial Electronics, 67(3): 2337-2347, 2020.

[9] S. Liu, Z. Hou, Y. Guo, and L. Guo, A novel modified robust model-free adaptive control method for a class of nonlinear systems with time delay, in IEEE 8th Data Driven Control and Learning Systems Conference, 2019: 1329-1334.

[10] Y. Li, W. Cao, W. Hu, Y. Xiong, and M. Wu, Incipient fault detection for geological drilling processes using multivariate generalized Gaussian distributions and Kullback-Leibler divergence, Control Engineering Practice, 2021, doi.org/10.1016/j.conengprac.2021.104937.

[11] Y. Cao, X. Yuan, Y. Wang, and W. Gui, Hierarchical hybrid distributed PCA for plant-wide monitoring of chemical processes, Control Engineering Practice, 2021, doi.org/10.1016/j.conengprac.2021.104784.

[12] A. Widodoa and B. S. Yang, Support vector machine in machine condition monitoring and fault diagnosis, Mechanical Systems and Signal Processing, 21(6): 2560-2574, 2007.

[13] H. Luo, K. Li, M. Hao, S. Yin, and O. Kaynak, A data-driven process monitoring approach with disturbance decoupling, in IEEE 7th Data Driven Control and Learning Systems Conference, 2018: 569-574.

[14] Y. Zhao, C. Zhao, and Y. Sun, Dynamic multivariate alarm threshold optimization for nonstationary processes subject to varying conditions, in IEEE 10th Data Driven Control and Learning Systems Conference, 2021: 836-841.

[15] Z. Zhang, J. Zhu, and Z. Ge, Industrial process modeling and fault detection with recurrent Kalman variational autoencoder, in IEEE 9th Data Driven Control and Learning Systems Conference, 2020: 1370-1376.

[16] P. V. Overschee and B. L. Moor, Subspace Identification for Linear Systems: Theory—Implementation—Applications. Norwell: Springer Science & Business Media, 2012, chapter 5.

[17] M. Döhler and L. Mevel, Subspace-based fault detection robust to changes in the noise covariances, Automatica, 49(9): 2734-2743, 2013.

[18] M. Basseville, M. Abdellahian, and A. Benveniste, Subspace-based fault detection algorithms for vibration monitoring, Automatica, 36(1): 101-109, 2000.

[19] H. Chen and B. Jiang, A review of fault detection and diagnosis for the traction system in high-speed trains, IEEE Transactions on Intelligent Transportation Systems, 21(2): 450-465, 2020.

[20] H. Chen, Z. Chai, O. Dogru, B. Jiang, and B. Huang, Data-driven designs of fault detection systems via neural networks-aided learning, IEEE Transactions on Neural Networks and Learning Systems, 2021, doi:10.1109/TNNLS.2021.3071292.

[21] J. Yang, K. Zhou, Y. Li, and Z. Liu, Generalized out-of-distribution detection: A survey, ArXiv Preprint, 2021, arXiv:2110.11334.

[22] H. Chen, Z. Liu, C. Alippi, B. Huang, and D. Liu, Explainable intelligent fault diagnosis for nonlinear systems: From unsupervised to supervised Learning, IEEE Transactions on Neural Networks and Learning Systems, 2022, to be submitted.

[23] J. Schmidhuber, Deep learning in neural networks: An overview, Neural Networks, 61: 85-117, 2015.

[24] Z. Pan, H. Chen, Y. Wang, B. Huang, and W. Gui, A new perspective on AE- and VAE-based process monitoring, IEEE Transactions on Cybernetics, to be accepted.

[25] K. E. S. Pilario and Y. Cao, Canonical variate dissimilarity analysis for process incipient fault detection, IEEE Transactions on Industrial Informatics, 14(12): 5308-5315, 2018.