Fault Diagnosis for A Class of Nonlinear Systems Based on Reinforcement Learning and Deterministic Learning

Zejian Zhu
Weiming Wu
Tianrui Chen
Cong Wang ( wangcong@sdu.edu.cn )

Research Article

Keywords: Reinforcement learning, Deterministic learning, Fault diagnosis, Nonlinear systems

Posted Date: August 8th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1930103/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License
Fault Diagnosis for A Class of Nonlinear Systems Based on Reinforcement Learning and Deterministic Learning

Zejian Zhu\textsuperscript{1}, Weiming Wu\textsuperscript{2}, Tianrui Chen\textsuperscript{2} and Cong Wang\textsuperscript{2*}

\textsuperscript{1}School of Automation Science and Engineering, South China University of Technology, Guangzhou, 510641, P. R. China.
\textsuperscript{2}School of Control Science and Engineering, Shandong University, Jinan, 250061, P. R. China.

*Corresponding author(s). E-mail(s): wangcong@sdu.edu.cn; Contributing authors: zzzhuzejian@hotmail.com; auwuweiming@163.com; trchen@sdu.edu.cn;

Abstract

In this paper a novel fault diagnosis (FD) approach combing the reinforcement learning (RL) and the deterministic learning theory (DLT) is proposed for a class of discrete-time nonlinear system with unknown dynamics. First, a bank of DLT-based dynamical neural network (NN) identifiers are utilized to achieve locally-accurate approximations of the unknown system dynamics along the normal and fault trajectories. Based on this, a novel feature learning method combing the RL with DLT is proposed to further adapt the NN weights to extract discriminative features. The extracted features are represented by constant NNs. Finally, constant NN-based dynamical estimators are constructed to achieve rapid FD. The novelties of the proposed methods are: 1) according to the DLT, the exponential convergence of the NN weights can be rigorously analysed based on the Lyapunov stability theory; 2) a new class of strategic utility function is designed based on the concept of the synchronization error in DLT-based dynamical pattern recognition approach, which is different from other RL-based FD techniques that penalise the future wrong FD decisions. Simulation results shows the practical significance of the proposed FD method.

Keywords: Reinforcement learning, Deterministic learning, Fault diagnosis, Nonlinear systems
1 Introduction

Reinforcement learning (RL) is a mathematical framework for solving the sequential decision-making (e.g., control and pattern recognition) problems in the dynamical environment [1]-[3]. In RL, an agent is employed to interact with the dynamical environment and learn the sequential decision-making actions that optimize the long-term reward function which indicates the goodness of the selected actions [4]. In recent years, there have been increasing research activities in the applications of RL on the fault diagnosis (FD) [5]-[13]. RL-based FD approaches can be mainly divided into two categories: the model-free and the model-based methods. In model-free methods there is not a model and thus the rewards and the optimal solutions are derived from the end-to-end based trial-and-error approach [14] while in model-based RL there exists a physical or learned dynamical model which is employed for the derivation of the rewards and optimal actions [15] [2].

Approximate/Adaptive Dynamic Programming (ADP) is one of the most popular model-based RL approaches to search for the optimal solution for a deterministic dynamical model [15][16]. The actor-critic structure is one of the most important component in the ADP approach [17]. In actor-critic structure, an action neural network (NN) derives a policy to interact with the environment. The strategic utility function is defined as the long-term system performance, and is employed to evaluate the action NN. The critic NN is applied to approximate the strategic utility function, that is, to estimate the long-term system performance [18][19]. Interesting results about the ADP/actor-critic methods for the dynamical system have been reported in the recent literatures, and have achieved excellent performance in FD [10]-[13]. For example, in [10], inspired from the ADP techniques, an active fault detector is designed for a class of discrete-time system, where a reward function is set to impose the penalty of the wrong fault detection decisions. In [11], an ADP-based active fault detector is employed for stochastic linear Markovian switching systems over an infinite time horizon to improve the discrimination between models. In [12], an ADP-based FD scheme is presented to detect and isolate faults in an aircraft jet engine.

So far, a great deal of progress has been made for ADP-based FD. However, only limited success has been reported in the literature for the rigorously analysis of the convergence to the optimal solution or the ideal parameters. Assuring convergence is a key open problem in RL [20]-[22]. In ADP-based FD methods, the convergences of the critic and action NN weights mean that the agents possess the ability to accurately estimate the long-term diagnosis performance and find the optimal FD scheme to obtain the optimal FD performance. Thus, it is significant to launch a major study on the parameter convergence analysis in the ADP-based FD algorithm. To guarantee the exponential convergence to the optimal solution or the ideal parameters, the persistent excitation (PE) condition is required [23]-[25]. For nonlinear dynamical systems, the PE condition is very difficult to characterize and usually cannot be verified a priori [26].
Recently, a deterministic learning theory (DLT) aiming at solving the problem of learning the unknown system dynamics for nonlinear dynamical system was proposed [27][28]. It stated that the partial PE condition for RBFNN regression subvector was satisfied almost always for a given period NN input orbit [28]. Due to the guaranteed satisfaction of partial PE condition, the exponential convergence of NN weights was achieved, and thus led to the locally-accurate NN approximation of the nonlinear system dynamics along the NN input trajectory. Based on the DLT, a rapid dynamical pattern recognition (DPR) approach was proposed to indicate that a time-varying dynamical pattern could be represented in a time-invariant and spatially distributed form, and rapid DPR can be achieved based on the difference between the system dynamics of the test and training patterns [29]. Utilizing the rapid DPR approach, in previous works, we have given analytical results showing that the obtained knowledge of the fault be utilized to achieve rapid FD scheme [30], with applications on the early detection of rotating stall [31], detection of heart valve disorders from phonocardiogram signals and so on [32].

In this paper, for the problem of the convergence analysis in the ADP-based FD, a rapid FD approach is proposed for a class of discrete-time nonlinear dynamical system by combing the DLT with ADP. The proposed approach consists the following three steps.

First, based on the DLT, locally-accurate approximations of the unknown system dynamics along the normal and fault patterns are achieved. The obtained dynamical features are saved in the form of the constant NNs. Notice that constructing an accurate dynamical model is the basis of a high performance RL method.

Second, a novel framework combing the RL and the DLT is proposed to further adapt the NN weights saved in the previous step, with the objective to propose a robust FD method whose parameter convergence can be rigorously analysed. To achieve this,

- A new strategic utility function is introduced as the long-term FD performance measure based on the DPR approach derived by the DLT. The strategic utility function is defined as the distance of the synchronization errors/residuals between the normal and fault patterns. Notice that the synchronization errors are the external reflection of the extracted dynamical features [29]. By minimizing the strategic utility function, the difference of the extracted dynamical features between the normal and fault patterns have been expanded. Thus discriminative features are extracted. Extracting discriminative features is the key of achieving high FD performance [38].
- The NN weight updating law is designed using the deterministic learning method to minimize the strategic utility function, and the exponential convergences of the critic and action NN weights are rigorously analysed based on the Lyapunov stability theory. The problems of estimating the strategic utility function and learning the optimal FD policy are turned into the stability issue of a class of discrete-time linear time-vary (LTV) systems.
Fault Diagnosis Based on Reinforcement Learning and Deterministic Learning

According to the DLT, when the NN input trajectory is a period or recurrent orbit, the Gaussian RBFNN regression subvector along the trajectory satisfies the PE condition, which leads to the exponential convergences of the critic and action NN weights. The learned knowledge is stored and represented in the form of the constant NNs.

Finally, by utilizing the learned knowledge represented by constant NNs, experience-based dynamical estimators are constructed to achieve rapid FD. The main contributions of this paper are highlighted as follows.

1) Compared with the existing ADP-based FD methods (e.g., [11][12]), the proposed method enables the agents to accurately estimate the long-term FD performance measure and rapidly search for the optimal action. This thanks to the satisfaction of the partial PE condition. Based on the satisfaction of the partial PE condition, the exponential convergences of the critic and action NN weights are rigorously analysed by using the Lyapunov direct method.

2) Compared with the DLT-based FD approaches (see [30]), the method proposed in this paper not only consider the approximation of the unknown system dynamics, but also the minimization of the long-term FD performance, and thus the FD performance is improved. The experiment results show that proposed method is able to improve the robustness of the DLT-based FD algorithm.

3) In the existing RL-based FD methods (e.g., [6][7][10], etc), the strategic utility function is defined as the penalty on the future wrong FD decisions, thus the agent learns a task-specific action from the penalty on the incorrect FD results. Compared with them, the strategic utility function presented in this paper is defined as a decreasing function of the distance of the normal and fault residuals. By minimizing the strategic utility function (thus to maximize the distance between the normal and fault patterns in the feature space), the agent is able to directly learn a more discriminative fault feature from the sampling sequences which contains more sufficient dynamical information. Such a kind of learning mechanism enables the agents to learn not only the similar dynamical behavior of the patterns belonging to the same class, but also capture the difference between the dynamical behavior of the normal and fault patterns.

The rest of the paper is organized as follows. Problem formulations and some preliminaries are given in Section II. The main results, including the proposed FD approach and the stability results, are presented in Section III. Simulation results are given in Section IV. The conclusions are drawn in Section V.
2 Problem Formulation and Preliminaries

2.1 Problem Formulation

Consider the FD problem for the following continuous nonlinear dynamical system given by

\[ \dot{X} = F(X) + \beta(t - T_0)G(X; p), \quad X(0) = X_0 \] (1)

where \( X \in \mathbb{R}^n \) is the state vector of the system. \( F(x) = [f_1(x; p), f_2(x; p), \cdots, f_n(x; p)]^T \) is a smooth vector field. \( G(x; p) = [g_1(x; p), g_2(x; p), \cdots, g_n(x; p)]^T \) represents the change in the system dynamics due to a fault. \( p \) is the fault parameter. \( \beta(t - T_0) \) is a step function representing the time profiles of the faults, where \( T_0 \) denotes the time when the fault occurs, i.e.,

\[ \beta(t - T_0) \triangleq \text{diag}\{\beta_1(t - T_0), \beta_2(t - T_0), \cdots, \beta_n(t - T_0)\} \] (2)

Assumption 2.1 The trajectories generated by the dynamical system (1) remain uniformly bounded in a compact set, i.e., \( \varphi \in \Omega_X, \forall t > 0, X_0 \in \Omega_X \), where \( \Omega_X \in \mathbb{R}^n \) is a compact set. Moreover, the system trajectory is a period or recurrent orbit.

A simple forward Euler discretization of the system (1) is given by

\[
X(K + 1) = X(K) + T_s(F(X(K)) + \beta(K - K_0)G(X(K); p)) \quad X(1) = X_0
\] (4)

where \( X(K) = [x_1(K), x_2(K), \cdots, x_n(K)] \in \mathbb{R}^n \) represents the system state vector. \( T_s \) is the sampling period. Denote the sampling sequences derived from the system (4) as \( \phi \triangleq \{X(1), X(2), \cdots, X(L)\} \), where \( L \) denotes the length of the sampling sequence. \( \beta(K - K_0) \) is the step function with \( K_0 \) represents the time when the fault occurs. Define \( H(X(K)) = F(X(K)) + \beta(K - K_0)G(X(K); p) \) as the general fault function. When \( K < K_0 \), the system is working under the normal operation condition, then the sequence derived from the system (4) is represented as \( \phi^0 \), and referred as the normal pattern. When \( K \geq K_0 \), the system is faulty, the sequences derived from (4) with different fault parameters and different initial conditions are denoted as \( \phi^s(s = 1, 2, \cdots) \) and referred as the fault patterns.

Consider the training set \( \mathcal{R} \triangleq \{\phi^0_1, \phi^0_2, \cdots, \phi^{0Q}_1, \phi^{11}_1, \phi^{12}_1, \cdots, \phi^{1Q}_1, \cdots, \phi^{M1}_1, \phi^{M2}_1, \cdots, \phi^{MQ}_1\} \), which contains the sequences of the normal pattern and \( M \) classes of fault patterns, and each class of pattern contains \( Q \) sequences.
sampled from the dynamical system (4) with different initial conditions. \( \phi^q_r = \{X^r_q(1), X^r_q(2), \ldots, X^r_q(L)\} \) \( r \in \{1, 2, \ldots, M\}, q \in \{1, 2, \ldots, Q\} \)

represents the \( q \)th sequence belonging to the \( r \)th class of fault pattern. For simplification, we denote \( \phi^q_r(q \in \{1, 2, \ldots, M\}, q \in \{1, 2, \ldots, Q\}) \) as the sequence generated by the \( r \)th fault pattern, and denote \( \phi^0_j(j \in \{1, 2, \ldots, Q\}) \) as the normal pattern.

In this paper we consider the FD problem as a pattern recognition problem. Our objective is to design a pattern recognition-based FD approach such that when the system is under the normal operation condition, it can be recognized as the normal pattern, and when the system is faulty, it can be recognized as the fault pattern.

2.2 RBFNN and Partial PE Condition

2.2.1 Universal Approximation Property of RBFNNs

For any continuous nonlinear function \( f(Z) : \mathbb{R}^n \mapsto \mathbb{R} \), the following RBFNN is employed to approximate it, i.e.,

\[
    f_{nn}(Z) = \sum_{i=1}^{q} w_i s_i(Z) = W^T S(Z)
\]

(5)

where \( Z \in \mathbb{R}^n \) is the NN input vector with \( n \) being the dimension of the NN input. \( W = [w_1, w_2, \ldots, w_q]^T \) is the NN weight vector. \( q \) is the number of the NN nodes. \( S(Z) = [s_1(Z), s_2(Z), \ldots, s_q(Z)]^T \) is the basis function vector with \( s_i(Z) \) chosen as a Gaussian function. It has been proven in [35] that if the number of NN nodes is large enough, the RBFNN can approximate any continuous function \( f(Z) \) to any arbitrary error \( \epsilon^* \) as

\[
    f(Z) = W^* S(Z) + \epsilon(Z), \quad Z \in \Omega_Z
\]

(6)

where \( W^* \in \mathbb{R}^q \) is the ideal NN weight vector. \( \epsilon(Z) < \epsilon^* \) is the ideal NN approximation error.

2.2.2 Localized RBFNN and Partial PE Condition

The RBFNN possesses the spatially localized learning capabilities of representation [27], which means that for any continuous nonlinear function \( f(Z) \), it can be approximated utilizing a limited number of NN nodes located in a local region along the trajectory, i.e.,

\[
    f(Z) = W_{\zeta}^* S_{\zeta}(Z) + \epsilon_{\zeta}
\]

(7)

where \( S_{\zeta}(Z) = [s_{1\zeta}(Z), s_{2\zeta}(Z), \ldots, s_{j\zeta}(Z)]^T \in \mathbb{R}^{j\zeta} \) denotes the Gaussian radial basis function consisting of the NN nodes located in the region along the NN input trajectory, with \( j_{\zeta} \) being the number of the corresponding NN nodes. \( \epsilon_{\zeta} \) is the NN approximation error using the localized RBFNN.
To show the PE condition of the RBFNN, the following definition was presented in [36]:

**Definition 1** Consider a uniform-bounded, vector-valued, discrete-time signal \( S(Z(K)) : R^n \mapsto R^q \). If there exist positive constants \( \alpha \) and \( K' \) such that

\[
\sum_{i=K}^{K+K'-1} S(Z(i))S(Z(i))^T \geq \alpha I, \quad K \geq 0
\]

where \( I \) is an identity matrix. Then \( S(Z(K)) \) is said to be persistently exciting.

**Lemma 2.1** [28] Consider any period or recurrent sequence \( Z(K) \). Assume that \( Z(K) \) is a discrete map from \([0, \infty)\) to a compact set \( \Omega_Z \subset R^q \). For the RBFNN with centers evenly placed on a regular lattice (large enough to cover the compact set \( \Omega_Z \)), \( S_{\zeta}(Z(K)) \) is persistently exciting almost always.

### 2.3 DLT-based FD

The DLT-based FD consists of two steps: the DLT-based system identification and the DPR-based FD. In the system identification phase, a bank of identifiers, including the normal identifiers and the fault identifiers are employed to achieve locally-accurate approximation of the system dynamics along the normal and fault patterns, respectively. The learned knowledge of the system dynamics are stored in the constant NNs. In the DPR-based FD phase, a set of constant NN-based dynamical estimators are utilized to achieve rapid FD based on the minimum residual principle. In the following we will demonstrate how these two steps are implemented.

#### 2.3.1 DLT-based System Identification

Consider the \( i \)th \((i = 1, 2, \cdots, n)\) subvector of the system (4)

\[
x_i^{s_j}(K+1) = x_i^{s_j}(K) + T_s h_i^{s_j}(X^{s_j}(K))
\]

where \( X^{s_j}(K) = [x_1^{s_j}(K), x_2^{s_j}(K), \cdots, x_n^{s_j}(K)]^T \) (\( s \in \{0, 1, 2, \cdots, M\}, j \in \{1, 2, \cdots, Q\} \)). When \( s \neq 0 \), \( X^{s_j} \) denotes the data of the \( s_j \)th fault pattern while when \( s = 0 \), \( X^{s_j} \) denotes the data of the normal pattern. \( h_i^{s_j}(X^{s_j}(K)) : R^n \mapsto R \) represents the unknown general fault function.

In the following, we will present the DLT-based method to achieve accurate NN approximation of the general fault function along the sequence \( \varphi^{s_j} \).

First, a dynamical RBFNN identifier is constructed as follows:

\[
\hat{x}_i^{s_j}(K+1) = \hat{x}_i^{s_j}(K) + (\alpha - 1)(\hat{x}_i^{s_j}(K) - x_i^{s_j}(K)) + T_s \hat{W}_i^{s_j}S(X^{s_j}(K)), \quad \forall s = 0, 1, 2, \cdots, Q
\]
when \( s \neq 0 \), the identifier (10) represents the fault identifier, and when \( s = 0 \), the identifier (10) represents the normal identifier.

Compare (10) with (9), and define \( \tilde{x}_i^{s_j}(K + 1) = x_i^{s_j}(K + 1) - x_i^{s_j}(K + 1), \) \( \tilde{W}_{\zeta_i}^{s_j} = \tilde{W}_{\zeta_i}^{s_j} - W_{\zeta_i}^{s_j} \). Using the spatially localized learning property of RBFNN, we have

\[
\tilde{x}_i^{s_j}(K + 1) = \alpha_a \tilde{x}_i^{s_j}(K) + T_s \tilde{W}_{\zeta_i}^{s_j}(K) S_\zeta(X^{s_j}(K)) + T_s \epsilon_{\zeta_i}(K) \tag{11}
\]

where \( \tilde{x}_i^{s_j} \in R (s \in \{0, 1, 2, \ldots, M\}, j \in \{1, 2, \ldots, Q\}) \) denotes the state vector of the RBFNN identifier. \( X^{s_j}(K) = [x_1^{s_j}(K), x_2^{s_j}(K), \ldots, x_n^{s_j}(K)]^T \in R^n \) is the state vector of the dynamical system under the \( s_j \)th pattern. \( \tilde{W}_{\zeta_i}^{s_j}(K) S(X(K)) \) is the RBFNN utilized to approximate \( h_i^{s_j}(X(K)) \).

The NN weights are updated by using the Lyapunov-based learning law:

\[
\tilde{W}_{\zeta_i}^{s_j}(K + 1) = \tilde{W}_{\zeta_i}^{s_j}(K) - T_s \gamma S(X(K)) \tilde{x}_i^{s_j}(K + 1) \tag{12}
\]

Utilizing the spatially localized learning property of RBFNN, the NN weight estimation error system can be expressed as

\[
\tilde{W}_{\zeta_i}^{s_j}(K + 1) = \tilde{W}_{\zeta_i}^{s_j}(K) - T_s \gamma S_\zeta(X^{s_j}(K)) \tilde{x}_i^{s_j}(K + 1) \tag{13}
\]

where \( \tilde{W}_{\zeta_i}^{s_j}(K) = \tilde{W}_{\zeta_i}^{s_j}(K) - W_{\zeta_i}^{s_j} \). Substitute (11) into (13), then we have

\[
\tilde{W}_{\zeta_i}^{s_j}(K + 1) = -T_s \alpha_a \gamma S_\zeta(S_\zeta(K)) \tilde{x}_i^{s_j}(K) + (I - T_s^2 \gamma S_\zeta(K) S_\zeta^T(K)) \tilde{W}_{\zeta_i}^{s_j}(K) - T_s^2 \gamma S_\zeta(K) \epsilon_{\zeta_i}(K) \tag{14}
\]

Then the dynamical error systems (11) and (13) are expressed in the following LTV form:

\[
\begin{bmatrix}
\tilde{x}_i^{s_j}(K + 1) \\
\tilde{W}_{\zeta_i}^{s_j}(K + 1)
\end{bmatrix} =
\begin{bmatrix}
\alpha_a & T_s S_\zeta^T(K) \\
-T_s \alpha_a \gamma S_\zeta(K) I - T_s^2 \gamma S_\zeta(K) S_\zeta^T(K)
\end{bmatrix} \times
\begin{bmatrix}
\tilde{x}_i^{s_j}(K) \\
\tilde{W}_{\zeta_i}^{s_j}(K)
\end{bmatrix} +
\begin{bmatrix}
T_s \epsilon_{\zeta_i}(K) \\
-T_s^2 \gamma S_\zeta(K) \epsilon_{\zeta_i}(K)
\end{bmatrix} \tag{15}
\]

Since the trajectory of \( X^{s_j}(K) \) is a recurrent orbit, according to Lemma 2.1, \( S_\zeta(X^{s_j}(K)) \) is persistently exciting. According to [37, Theorem 1], the exponential stability of \( (\tilde{x}_i^{s_j}, \tilde{W}_{\zeta_i}^{s_j}) = 0 \) for the nominal part of the system (15) is achieved. Since the terms \( T_s \epsilon_{\zeta_i}(X^{s_j}(K)) \) and \( T_s^2 \gamma S_\zeta(X^{s_j}(K)) \epsilon_{\zeta_i}(X^{s_j}(K)) \) are bounded, we obtain that \( \tilde{W}_{\zeta_i}^{s_j} \) converges exponentially to the small neighborhood of zero, with the size of neighborhood determined by \( T_s \) and \( \epsilon_\zeta \). This means that accurate identification of the unknown system dynamics along the
sequence $\phi^{s_j}$ is achieved. Then a constant NN can be utilized to obtain the spatially distributed representation of the general fault function $h_i^{s_j}(X^{s_j})$, i.e.,

$$h_i^{s_j}(X^{s_j}(K)) = \bar{W}^{s_j}_i S(X^{s_j}(K)) + \bar{\epsilon}^{s_j}(K)$$  \hspace{1cm} (16)

where

$$\bar{W}^{s_j}_i = \frac{1}{K_b - K_a + 1} \sum_{K=K_a}^{K_b} \hat{W}^{s_j}_i(K)$$  \hspace{1cm} (17)

$\bar{\epsilon}^{s_j}(K)$ is the NN approximation error. $[K_a, K_b]$ is the time segment after the NN weight convergence. When $s \neq 0$, (17) represents the fault NN weights, and when $s = 0$, (17) represents the normal NN weights. In this way, the knowledge of the normal and fault patterns are stored in the normal and fault NNs.

### 2.3.2 DLT-based Rapid FD

Consider the test pattern $\phi^t$. Based on the acquired knowledge of $h_i^{s_j}(X^{s_j}(K))$, a rapid FD mechanism is demonstrated as follows. First, a bank of normal and fault NN-based dynamical estimators are constructed as follows:

$$\hat{x}^{s_j}_i(K + 1) = \hat{x}^{s_j}_i(K) + (\alpha - 1)(\hat{x}^{s_j}_i(K) - x^t_1(K))$$

$$+ T_s \bar{W}^{s_j}_i S_i(X^t(K))$$  \hspace{1cm} (18)

where $\hat{x}^{s_j}_i \in R$ ($s \in \{0, 1, 2, \cdots, M\}$, $j \in \{1, 2, \cdots, Q\}$) denotes the state of the estimator. $X^t(K) = [x^t_1(K), x^t_2(K), \cdots, x^t_n(K)]^T \in R^n$ is the state vector of the test pattern. $\bar{W}^{s_j}_i S_i(X^t(K))$ is the constant RBFNN utilized to approximate the system dynamics of the test pattern. An average $L_1$ norm of the residual is generated and utilized for FD scheme:

$$\|\bar{x}^{s_j}_i(K)\|_1 = \frac{1}{T} \sum_{k=K-T}^{K} |\bar{x}^{s_j}_i(k)|$$  \hspace{1cm} (19)

where $T$ is the length of the average $L_1$ norm. $\bar{x}^{s_j}_i(k) = \hat{x}^{s_j}_i(k) - x^t_1(k)$ is the synchronization error/residual.

The DLT-based FD scheme is given based on the minimum residual principle. Compare $\|\bar{x}^{t_j}_i(K)\|_1$ with $\|\bar{x}^{0_j}_i(K)\|_1$, $r \in \{1, 2, \cdots, M\}$, $j \in \{1, 2, \cdots, Q\}$, $q \in \{1, 2, \cdots, Q\}$. If, for some $r$, there exists some finite time $K_d$ such that $\|\bar{x}^{t_j}_i(K)\|_1 < \|\bar{x}^{0_j}_i(K)\|_1$, then it is deduced that the fault occurs.

**Remark 1** The main idea of the DLT-based FD scheme is as follows. It is based on the minimum residual principle. The fault detection signal is given when the fault residual $\|\bar{x}^{t_j}_i(K)\|_1$ is less than the normal residual $\|\bar{x}^{0_j}_i(K)\|_1$. This method is employed in FD for oscillation systems such as the early detection of the rotating
stall [31]. Since according to [29][30], the synchronization errors/residuals are the external reflection of the difference of the extracted features between two dynamical systems, we wonder how this concept of the synchronization errors/residuals can be employed to extract discriminative features, which is shown to be crucial to improve the robustness of the FD algorithm. A natural idea is to make $\|\tilde{x}_r(K)\|_1$ far more than $\|\tilde{x}_i(K)\|_1$ when the system is under the normal operation condition. In the next section, we will demonstrate that how to use the concept of synchronization errors/residuals to extract discriminative features by combing the RL and the DLT.

## 3 RL and DLT Based FD

In this section we demonstrate how to use the concept of the synchronization errors/residuals to design the RL and DLT-based FD framework. The proposed framework is shown Fig. 1. First, a DLT-based system identification method is utilized to achieve locally-accurate approximation of the unknown system dynamics along the normal and fault patterns as shown in section 2.3.1. The learned knowledge is stored in NNs. Second, we combine the DLT and RL methods to further adapt the NN weights with the objective of extracting discriminative features. The exponential convergences of the critic and action NN weights are proved using the Lyapunov direct method. Finally, the converged NN weights are stored in constant NNs and a bank of experience-based dynamical estimators are utilized to achieve rapid FD as shown in section 2.3.2.
In this section we mainly demonstrate how the second step is implemented, that is, how to combine the RL and the DLT to further adapt the NN weights. The proposed method is based on the actor-critic structure as shown in Fig. 2. The distinguishes of the proposed method in this section are the following two points. The first point is that the convergence of the proposed approach can be guaranteed and rigorously analysed, which enables the agents learn an optimal policy quickly. The second point is that a novel strategic utility function is defined using the concept of the synchronization errors/residuals as shown in section 2.3.2, aiming at extracting discriminative features.

The proposed method using a DLT-embedded actor-critic structure. Each identifier contains a critic NN and an action NN. For the normal identifier, the steps are shown as the red line in Fig. 2. We substitute the fault pattern into the normal identifier and fault identifier respectively. Since we our objective is to extract discriminative features, the generated normal and fault residuals are utilized to construct the strategic utility function such that when the test pattern is similar to the normal pattern, the generated $L_1$ norm of the fault residual should be far more than the normal residual. By minimizing the strategic utility function, the distance of the normal and fault patterns in the feature space is expanded, and thus the agents are able to learn a discriminative fault feature. The critic NN is utilized to approximate the strategic utility function. The action NN, whose initial NN weight vector is set as the saved constant NN weight vector obtained from the DLT-based system identification.
method mentioned above, is updated to minimize the strategic utility function. According to the DLT, the partial PE condition of the RBFNN regression subvector is satisfied, then the exponential convergence of the critic and action NN is rigorously analysed using the Lyapunov stability theory. The design of the fault identifier is similar to the normal identifier, which is shown as the green line in Fig. 2.

3.1 Strategic Utility Function

In the following we will show how the strategic utility function for the $j$th normal identifier is designed. 

First, we substitute the sequence $\phi^{r_q} = \{X^{r_q}(1), X^{r_q}(2), \cdots, X^{r_q}(L)\}$ \((\forall r \in \{1, 2, \cdots, M\}, q \in \{1, 2, \cdots, Q\})\) into the $j$th normal identifier, i.e.,

$$
\hat{x}_i^{(0_j,r_q)}(K + 1) = \hat{x}_i^{(0_j,r_q)}(K) + (\alpha_a - 1)(\hat{x}_i^{(0_j,r_q)}(K) - x_i^{r_q}(K)) \\
+ T_s \hat{W}_i^{(0_j,r_q)}(K)S(X_f^{r_q}(K))
$$

(20)

$x_i^{r_q}(K)$ is the $i$th subvector of $X^{r_q}(K)$ \((K = 1, 2, \cdots, L)\). $\hat{x}_i^{(0_j,r_q)}(K)$ is the state vector of the normal identifier in the case when substituting the sequence $\phi^{r_q}$ into it. $\hat{W}_i^{(0_j,r_q)}$ is the NN weight vector of the $j$th normal action NN when substituting the $r_q$th sequence into it.

Second, we substitute the $r_q$th fault sequence into the $r_k$th fault identifier, i.e.,

$$
\hat{x}_i^{(r_k,r_q)}(K + 1) = \hat{x}_i^{(r_k,r_q)}(K) + (\alpha_a - 1)(\hat{x}_i^{(r_k,r_q)}(K) - x_i^{r_q}(K)) \\
+ T_s \hat{W}_i^{(r_k,r_q)}(K)S(X_f^{r_q}(K)), \forall k = 1, 2, \cdots, Q
$$

(21)

$\hat{x}_i^{(r_k,r_q)}$ is the $i$th state vector in the case when substituting the $r_q$th sequence into the $r_k$th fault identifier. $\hat{W}_i^{r_k}$ is the NN weight vector of the $r_k$th fault action NN. The initial values are set as $\hat{W}_i^{(0_j,r_q)}(1) = \hat{W}_i^{0_j}$, $\hat{W}_i^{(r_k,r_q)}(1) = \hat{W}_i^{r_k}$, where $\hat{W}_i^{0_j}$ and $\hat{W}_i^{r_k}$ are stated in (17). For simplification, $(\cdot)^{(0),r_q}$ and $(\cdot)^{(r_k,r_q)}$ are denoted by $(\cdot)^0$ and $(\cdot)^r$, respectively.

The normal identifier is designed with the aim that when substituting the $r_q$th fault sequence into the normal and fault identifiers, the $L_p$ norm of the normal residual is far more than the $r_k$th fault residual, $\forall k \in \{1, 2, \cdots, Q\}$. This objective can be implemented by the optimization of the distance between the normal and fault patterns in the feature space, i.e.,

$$
R_i^0(X^{r_q}(K)) = \beta(h(X^{r_q}(K))), \quad \forall k = 1, 2, \cdots, Q
$$

(22)

where $\beta(\cdot)$ is a decreasing function of $h(X^{r_q}(K))$, i.e., $\beta(h(X^{r_q}(K))) = \lambda e^{-h(X^{r_q}(K))}$. The nonlinear term $h(X^{r_q}(K))$ can be designed as

$$
h(X^{r_q}(K)) = \|\hat{W}_i^0(K)S(X_f^{r_q}(K))\|_p - \|\hat{W}_i^r(K)S(X_f^{r_q}(K))\|_p
$$

(23)
or
\[
h(X^{rs}(K)) = \|\tilde{x}_i^0(K)\|_p - \|\tilde{x}_i^r(K)\|_p
\]  
(24)

where \(\|\cdot\|_p\) denotes the averaged \(L_p\) norm.

The utility function derived by \(R_i^0(X^{rs}(K))\) is designed as the function of \(X^{rs}(K)\), i.e.,
\[
P_i^0(X^{rs}(K)) = 1 - \frac{1}{1 + R_i^0(X^{rs}(K))}, \forall k = 1, 2, \cdots, Q
\]  
(25)

Inspired from [34], the long-term FD performance measure or the strategic utility function derived by \(P_i^0(X^{rs}(K))\) is designed as follows:
\[
J_i^0(X^{rs}(K)) = \sum_{w=K+1}^{\infty} \alpha^{N-w+K+1}P_i^0(X^{rs}(w)), \forall k = 1, 2, \cdots, Q
\]  
(26)

where \(0 < \alpha < 1\), \(N\) is the horizon. Similar to [33][34], the equation (26) can be expressed as
\[
J_i^0(X^{rs}(K)) = \min_{W_{ic}(K)} \{\alpha J_i^0(K - 1) - \alpha^{N+1}P_i^0(K)\}, \forall q = 1, 2, \cdots, Q
\]  
(27)

### 3.2 Critic NN

In this subsection, our objective is to design a critic NN \(\hat{J}_i^0(X_f(K)) \triangleq \hat{W}_{ic}^{0T}(K)S(X^{rs}(K))\) to approximate \(J_i^0(X_f(K))\). \(\hat{W}_{ic}^0(K) \in R^l\) is the estimated weight vector of the critic NN. \(l\) is the number of NN nodes. \(S(X^{rs}(K)) \in R^q\) is the Gaussian radial basis function vector with the input \(X^{rs}(r \in \{1, 2, \cdots, M\}, q \in \{1, 2, \cdots, Q\})\).

The prediction error for the critic NN derived by \(P_i(X^{rs}(K))\) is
\[
e_{ic}(K) = \hat{J}_i^0(K) - \alpha \hat{J}_i^0(K - 1) + \alpha^{N+1}P_i^0(K)
\]  
(28)

Similar to [33], the updating law of the critic NN is designed with the objective to minimize the quadratic objective function \(E_{ic}(K) = \frac{1}{2}e_{ic}^T(K)e_{ic}(K)\) based on the gradient descent, i.e.,
\[
\dot{\hat{W}}_{ic}^0(K + 1) = \hat{W}_{ic}^0(K) - \alpha_c S(X^{rs}(K))
\times (\hat{W}_{ic}^0(K)S(X^{rs}(K)) + \alpha^{N+1}P_i^0(X^{rs}(K))
- \alpha \hat{W}_{ic}^0(K - 1)S(X^{rs}(K - 1))
\]  
(29)

\(\alpha_c > 0\) is a design parameter.
3.3 Action NN

In this subsection, the action NN $\hat{W}_{ia}^{0T}S(X^{r_q}(K))$ is employed to minimize $R_i^0(X^{r_q})$ and $J_i^0(X^{r_q})$ along the trajectory of the sequence $X^{r_q}$. The initial value of the action NN weight vector is $\hat{W}_{i}^{0jT}$, which is given in (17). Define $\Upsilon_i^0(X^{r_q}(K)) \triangleq R_i(X^{r_q}(K)) + J_i^0(X^{r_q}(K))$ as a combined strategic utility function. For the design of the action NN weight updating law, along the sequence $\varphi^{r_q} = \{X^{r_q}(1), X^{r_q}(2), \cdots, X^{r_q}(L)\}$, first we consider the dynamical systems derived by the strategic utility function $\Upsilon_i^0(X^{r_q}(K))$ (denoted as $\Upsilon_i^0(K), K = 1, 2, \cdots, L$), i.e.,

$$\Upsilon_i^0(K + 1) = \Upsilon_i^0(K) + T_s h_\Upsilon(X^{r_q}(K))$$

$$= \alpha_\Upsilon \Upsilon_i^0(K) + T_s h_\Upsilon(X^{r_q}(K)) + (1 - \alpha_\Upsilon)\Upsilon_i^0(K)$$

(30)

where $\alpha_\Upsilon$ is a parameter which will be given later. $h_\Upsilon(\cdot)$ denotes the unknown nonlinear function which is relative to $X^{r_q}$.

From the viewpoint of the control theory, the problem of minimizing the combined strategic utility function $\Upsilon_i^0$ can be solved by stabilizing the systems (30). Here the action NN $\hat{W}_{ia}^{0T}(K)S(X^{r_q}(K))$ is employed. Since $\Upsilon_i^0(X^{r_q}(K))$ and $\hat{W}_{ia}^{0T}(K)S(X^{r_q}(K))$ are both the functions of $X^{r_q}(K)$, equations (30) can be written as

$$\Upsilon_i^0(K + 1) = \alpha_\Upsilon \Upsilon_i^0(K) + T_s [\hat{W}_{ia}^{0T}(K)S(K) - g_i(K)]$$

$$= \alpha_\Upsilon \Upsilon_i^0(K) + T_s \hat{W}_{ia}^{0T}(K)S(K)$$

(31)

where $g_i(K) = \frac{1}{T_s} (T_s \hat{W}_{ia}^{0T}(K)S(K) - T_s h_\Upsilon(X^{r_q}(K)) - (1 - \alpha_\Upsilon)\Upsilon_i^0(K))$.

For the action NN along the fault sequence $\varphi^{r_q}$, the Lyapunov-based learning law is given by

$$\hat{W}_{ia}^{0T}(K + 1) = \hat{W}_{ia}^{0T}(K) - T_s \gamma S(X^{r_q}(K))\Upsilon_i^0(K + 1)$$

(32)

where $\gamma > 0$ is the learning rate. The initial condition is given by $\hat{W}_{ia}^{0T}(1) = \hat{W}_{ia}^{0j}$, where $\hat{W}_{ia}^{0j}$ is given by (17).

**Remark 2** Notice that the equation (31) is not coded in the program since the action NN does not affect $\Upsilon_i^0(X^{r_q}(K))$ directly. Only the updating law (32) is implemented. Here, since $\Upsilon_i^0$ is the function of $X^{r_q}$, the system dynamics in equations (30) can be represented in the form of $\hat{W}_{ia}^{0T}(K)S(X^{r_q}(K)) - g_i(X^{r_q}(K))$ due to the universal approximation property of the RBFNN.

3.4 Stability

The purpose of this section is to establish our main results by using Lyapunov direct method. Before engaging in demonstrating the main theorem, we need to develop some mild assumptions and facts as follows:
Assumption 3.1 The norms of the ideal critic and action NN weight vectors $W_{ic}^{0*}$ and $W_{ia}^{0*}$ are bounded over a compact set $\Xi_1$ by known positive constants $\bar{W}_{ic}^{0*}$ and $\bar{W}_{ia}^{0*}$, respectively, i.e.,

$$\|W_{ic}^{0*}\| \leq \bar{W}_{ic}^{0*}, \quad \|W_{ia}^{0*}\| \leq \bar{W}_{ia}^{0*}$$ (33)

Assumption 3.2 The NN approximation errors $\epsilon_{ic}^0(K)$ and $\epsilon_{ia}^0(K)$ are bounded by known positive constants $\bar{\epsilon}_{ic}$ and $\bar{\epsilon}_{ia}$ over a compact set $\Xi_2$, i.e.,

$$\|\epsilon_{ic}^0(K)\| \leq \bar{\epsilon}_{ic}, \quad \|\epsilon_{ia}^0(K)\| \leq \bar{\epsilon}_{ia}$$ (34)

Assumption 3.3 The norm of the Gaussian function vector $S(X^{r_q}(K))$ is uniformly bounded by a positive constant $s$ over a compact set $\Xi_3$, i.e.,

$$\|S(X^{r_q}(K))\| \leq s$$ (35)

Theorem 1 Consider the dynamical system (9). Let Assumptions 1-4 hold. Let the weight updating laws for the critic NN and the action NN be (29) and (32), respectively. Then we obtain that i) $\Upsilon_i^0(K)$, $J_i^0(K-1)$, $\bar{W}_{ic}^0$ and $\bar{W}_{ia}^0$ are all uniformly ultimately bounded; ii) $\bar{W}_{\zeta ic}^0(K)$ and $\bar{W}_{\zeta ia}^0(K)$ are exponentially stable, provided that when the following assumptions are hold:

$$0 < \alpha_c \|S(X^{r_q}(K))\|^2 < 1$$

$$0 < \alpha_T < 1$$

$$\gamma < \min\{\frac{1 - \alpha_T^2}{T_s^2\alpha_T^2 + \alpha_T T_s}, \frac{1}{T_s^2\alpha_T + T_s}\}$$ (36)

where $T_s$ is the sampling period. $s$ is the upper bound of $S(X^{r_q}(K))$, which is given in Assumption 3.3. $\alpha_T$ is the parameter given in (31).

Proof i) The proof of the UUB property is based on the Lyapunov stability theory. The readers are referred to Appendix for more details.

ii) First, we consider the error system of $\bar{W}_{\zeta ic}^0$, i.e.,

$$\bar{W}_{\zeta ic}^0(K+1) = (I - \alpha_cS_\zeta(K)S_\zeta^T(K))\bar{W}_{\zeta ic}^0(K)$$

$$- \alpha_cS_\zeta(K) \times (\bar{W}_{\zeta ic}^{0T}(K)S_\zeta(K) + \alpha^{N+1}P_i^0(K) - \alpha J_i^0(K-1))$$ (37)

where $S_\zeta(\cdot) \in R^{2K}$ is a subvector of $S(\cdot)$, which consists of the neurons located in the region close to the sequence $\varphi(X^{r_q})$. $J_i^0(K-1) = \bar{W}_{\zeta ic}^{0T}(K-1)S_\zeta(K-1)$.

According to Lemma 2.1, the regression subvector $S_\zeta(\cdot)$ satisfies the PE condition. Then according to [43], the exponential stability for the nominal part of the system (37) is achieved. Considering the assumptions 1-4, we have that $W_{\zeta ic}^{0T}(K)S_\zeta(K)$ is bounded, i.e., $\|W_{\zeta ic}^{0T}(K)S_\zeta(K)\| \leq \bar{W}_{\zeta ic}^{0T}s$. Since the UUB of $\bar{W}_{\zeta ic}^{0T}(K-1)S(X^{r_q}(K-1))$ has been proved in Theorem 1(i), which implies the UUB of $J_i^0(K-1)$. Thus, we obtain that $\bar{W}_{\zeta ic}^0$ converges exponentially to the neighborhood of zero, with the size of neighborhood determined by the terms $W_{\zeta ic}^{0T}(K)S_\zeta(K) + \alpha^{N+1}P_i^0(K) - \alpha J_i^0(K-1)$.

For the action NN weight estimation error vector $\bar{W}_{\zeta ia}^0$, we have

$$\left[ \begin{array}{c} \Upsilon_i^0(K+1) \\ \bar{W}_{\zeta ia}^0(K+1) \end{array} \right] = \left[ \begin{array}{cc} \alpha_R & T_sS_\zeta(K) \\ -T_s\alpha_R\gamma S_\zeta(K) & I - T_s^2\gamma S_\zeta(K)S_\zeta^T(K) \end{array} \right] \times \left[ \begin{array}{c} \Upsilon_i^0(K) \\ \bar{W}_{\zeta ia}^0(K) \end{array} \right]$$
\[
\begin{align*}
&\quad + \left[ T_s \epsilon_\zeta(K) - T_s^2 \gamma S_\zeta(K) \zeta^0(K) \right] \\
&= T_s \epsilon_\zeta(K) - T_s^2 \gamma S_\zeta(K) \zeta^0(K) \\
&= T_s \epsilon_\zeta(K) - T_s^2 \gamma S_\zeta(K) \zeta^0(K). 
\end{align*}
\] (38)

Since \( R_i(K) \) and \( \tilde{J}^0_i(K) \) have been proven to be UUB in Theorem 1, which leads to the UUB of \( \Upsilon_0^i \). According to , the exponential stability of \( \Upsilon_0^i, W_\zeta^0 i_a \) is achieved. Since the terms \( T_s \epsilon_\zeta(X^{r_\zeta}(K)) \) and \( T_s^2 \gamma S_\zeta(X^{r_\zeta}(K)) \zeta^0(X^{r_\zeta}(K)) \) are bounded, we obtain that \( \tilde{W}^0 _\zeta i_a \) converges to the small neighborhood of zero, with the size of neighborhood determined by \( T_s \) and \( \epsilon_\zeta \). The exponential stability of \( \tilde{W}^0 _\zeta i_a \) implies that \( \hat{W}^0 _\zeta i_a \) converges to the small neighborhood of \( W^\ast_\zeta i_a \). This ends the proof. □

Due to the convergence of \( \tilde{W}^{(0_j,r_k)}_{ic} \) and \( \tilde{W}^{(0_j,r_k)}_{ia} \), then the following equality is utilized to combine the knowledge obtained along the normal and faulty trajectory,

\[
\begin{align*}
\tilde{W}^{0_j}_{ic} &= \frac{\sum_{k=1}^{Q} \tilde{W}^{(0_j,r_k)}_{ic}}{Q} \\
\tilde{W}^{0_j}_{ia} &= \frac{\sum_{k=1}^{Q} \tilde{W}^{(0_j,r_k)}_{ia}}{Q} 
\end{align*}
\] (39)

where \( \tilde{W}^{(0_j,r_k)}_{ic} = \frac{1}{K_a-K_b+1} \sum_{w=K_b}^{K_a} \tilde{W}^{(0_j,r_k)}_{ic}(w) \), \( \tilde{W}^{(0_j,r_k)}_{ia} = \frac{1}{K_a-K_b+1} \sum_{w=K_b}^{K_a} \tilde{W}^{(0_j,r_k)}_{ia}(w) \). \([K_a,K_b]\) denotes the time segment after \( R_i \) converges to an acceptable value, which means that the distance between the normal pattern \( \phi^{0_j} \) and the fault pattern \( \phi^{r_\zeta} \) in the feature space is acceptable.

The design of the strategic utility function of the fault identifier is similar to the above, hence we do not give it here.

Remark 3 In this section our objective is to design a robust FD framework. To improve the robustness of the FD algorithm, a natural idea is to learn a discriminative fault feature, which is shown to be an important way to achieve high FD performance [38]. One powerful way to extract discriminative fault feature is to group the similar samples in the feature space while pushing dissimilar samples far apart from each other (see [39]-[42]), that is, to expand the distance between different class of dynamical patterns in the feature space.

Remark 4 In the existing RL-based FD techniques, the strategic utility function/reward function is designed based on the long-term FD performance measure such as the penalty on the future wrong fault detection decisions or the fault detection accuracy (see [7][10][11], etc). In this section, using the concept of the synchronization errors/residuals as shown in section 2.3.2, we regard the long-term FD performance as the distance between the normal and fault residuals, which is different from the evaluation on the FD results. Notice that in [29], the synchronization error (also named residual) is proved to be the external reflection of the extracted dynamic feature, and as shown in section 2.3.2, the faults are detected when the fault residuals
become smaller than the normal residuals. Thus, to avoid the false alarm, we expand the distance between the normal and fault residuals, and discriminative features are extracted to improve the robustness of the FD algorithm.

4 Simulation

In this section, we investigate the rotating stall detection problem for Mansoux model, aiming at pointing out the practical significance of the proposed FD method. Mansoux model is a high-order discrete model for analyzing the rotating stall inception and surge phenomenon in axial compressors. Rotating stall and surge are violent limit cycle-type oscillations in axial compressors. The FD experiment on this model is particularly important since it can describe the transient behavior of stall inception and coincide well with experimental results within a certain precision [44]. The Mansoux model is given as follows:

\[
E \dot{\phi} = -A \phi + \psi_{cd}(\phi) - T \dot{\psi}_p \\
\dot{\psi}_p = 1/(4l_cB^2)(S\phi - \Phi_T(\psi_p))
\]  

(40)

where \( \phi = [\phi_1, \phi_2, \ldots, \phi_M]^T \in R^M, M = 2N + 1 \) is the flow coefficient spaced around the annulus. \( M \) represents the number of sensors evenly placed around the annulus to describe the flow coefficient \( \phi \). The specific forms of matrices \( E \) and \( A \) are presented in [44].

The simulations are performed on an 18th-order model with \( N = 8 \), \( M = 17 \). The model parameters \( B \) and \( l_c \) are set the same as the Mansoux-C2 compressor dimensionless geometrical parameters [44]. The process of rotating stall is simulated through continuous adjustment of the exhaust characteristic parameter \( KT \) from 7 to 9.41. Partial sampled data of (40) in the time domain are shown in Fig. 3. As shown in Fig. 3, the rotating stall occurs at around \( T_{stall} = 350 \) rotor revolution. Our objective is to obtain a reliable and early stall warning signal. Thus, four training patterns, including three normal patterns and one stall inception pattern are selected. The selected three normal patterns are represented as \( D_{[220,234]}, D_{[235,250]}, D_{[250,260]} \), respectively. The selected fault pattern is represented as \( D_{[263,275]} \), where \( D_{[a,b]} \) represents the data sampled at rotor revolution segment \([a,b]\).

4.1 Identification of Selected Patterns

In this subsection, based on the DLT, the system dynamics under the normal and fault patterns are identified by selecting the data from probe 1 only. To achieve this, a discrete-time high gain observer is constructed to obtain the high dimensional system state trajectory and the DLT is utilized to achieve locally-accurate approximation of the system dynamics of the four selected patterns [45].

In the identification phase, the Gaussian RBFNNs which are applied to approximate the unknown system dynamics under the four different patterns
are constructed using 2500 nodes, with centers evenly placed on a regular lattice $[-11.5, 11.5] \times [-9.64, 9.64]$, and the width $\eta = 0.38$. The learning rate is set as $\gamma = 100$.

In Fig. 4, it is seen that the convergence of the NN weight vector $\hat{W}^f$ is achieved. The convergences of $\hat{W}^{n_i}(i = 1, 2, 3)$ are similar to $\hat{W}^f$, hence we do not given them here. After the convergence of NN weights, constant NN weight vectors are applied to store the learned knowledge, i.e.,

$$
\bar{W}^{n_i} = \frac{1}{K_b^{n_i} - K_a^{n_i} + 1} \sum_{K = K_a^{n_i}}^{K_b^{n_i}} \hat{W}^{n_i}(K), \quad i = 1, 2, 3
$$

$$
\bar{W}^f = \frac{1}{K_b^f - K_a^f + 1} \sum_{K = K_a^f}^{K_b^f} \hat{W}^f(K)
$$

(41)

where $\bar{W}^{n_i}$ denotes the constant NN weight vector of the $i$th normal pattern. $\bar{W}^f$ denotes the faulty constant NN weight vector. $[K_a^{n_i}, K_b^{n_i}]$ and $[K_a^f, K_b^f]$ are the corresponding time segments after the NN weight convergence.

### 4.2 Discriminative Feature Learning

In this subsection, based on the saved constant NN $\bar{W}_1$, we will combine the DLT with RL algorithm to further extract discriminative features from the selected four patterns. The initial values of $\bar{W}_a$ are set as $\bar{W}^{n_i}_a(1) = \bar{W}^{n_i}$ for $i \in \{1, 2, 3\}$, and $\bar{W}^f_a(1) = \bar{W}^f$. $\alpha_c$ in (27) is set as $\alpha_c = 0.9$.

The design of the reward function is set as follows. For $\bar{W}^{n_i}$, $i \in \{1, 2, 3\}$, we substitute the data sampled from the fault pattern into the $i$th normal identifier and the faulty identifier to derive the reward $R^{n_i}(X^f)$. The long-term reward function generated by the corresponding critic NN.
\( \hat{W}_f^T S(X_f) \) is denoted as \( \hat{J}^n_s(X_f) \). Denote the action NN weight vector updated by \( \Upsilon^n f(X_f) \) as \( \hat{W}_a^n \). After the convergence of \( \hat{W}_a^n \), the \( i \)th normal constant NN weight vector is saved as: \( \bar{W}_a^n = \frac{1}{K^b_i - K^a_i + 1} \sum_{K=K^a_i}^{K^b_i} \hat{W}_{n_i}(K) \), where \([K^a_i, K^b_i]\) is the time segment after the convergence of \( \hat{W}_a^n \).

For \( \hat{W}_a^n \), we substitute the data sampled from the normal pattern \( i(i \in \{1, 2, 3\}) \) into the fault pattern and the \( j(j \in \{1, 2, 3\}) \)th normal identifier to obtain the reward \( R(f,n_j)(X_{n_i}) \). The long-term reward function derived by the corresponding critic NN \( \hat{W}_c^{f,n_j} T S(X_{n_i}) \) is denoted as \( \hat{J}^{f,n_j}(X_{n_i}) \). Denote the action NN weight vector updated by \( \Upsilon^{f,n_j}(X_{n_i}) := R^{f,n_j}(X_{n_i}) + \hat{J}^{f,n_j}(X_{n_i}) \) as \( \hat{W}^{f,n_j}(X_{n_i}(K)) \). After the convergence of \( \hat{W}^{f,n_j} \), a combined constant faulty action NN weight vector \( \bar{W}^f(K) \) is set as \( \bar{W}^f(K) = \frac{1}{K^b_f - K^a_f + 1} \sum_{K=K^a_f}^{K^b_f} \hat{W}^{f,n_j}(X_{n_i}(K)) \). Figs. 5(a) and (b) shows the partial convergence of the critic NN weights \( \hat{W}_c^{f,n_1} \) and the action NN weights \( \hat{W}^{f,n_1} \). It is seen that partial NN weights convergence to relatively large values while the others remain in the small neighborhood of zero. This is consistent with the satisfaction of the partial PE condition. The convergences of the other critic and action NN weights are similar to the above, hence we do not give them here.

Fig. 6 shows the convergence of the combined long-term reward function \( \Upsilon^{f,n_1}(X_{n_2}) \). The convergence of \( \Upsilon^{f,n_1}(X_{n_2}) \) means that the distance between \( \hat{W}^{f,n_1} T S(X_{n_2}(K)) \) and \( \hat{W}^{(n_1.f)} S(X_{n_2}(K)) \) increases. This indicates that when the system works under the \( n_2 \)th normal operation condition, what the faulty NN and the normal NNs recall are mutually different, in other words, discriminative features have been extracted. The convergence of \( \Upsilon^{f,n_1}(X_{n_i}) \), \( i = 1, 3 \) is similar to the above, hence we do not give it here.
Fig. 5 Partial convergence of the NN weights using the DLT-RL based method. (a): Critic NN weights $\hat{W}_c^{(f,n_1)}$. (b): Action NN weights $\hat{W}_a^{(f,n_1)}$.

Fig. 6 Convergence of $\Upsilon^{(f,n_1)(X^{n_2})}$.

4.3 Detection of Rotating Stall

Two test systems are given as follows. The first test system is given by adding Gaussian noise to the training data, with the objective of verifying that discriminative features were extracted. For the second test system, a random parameter perturbation with the variance of 0.001 is added to the parameters $B$, $l_c$, matrices $A$ and $E$, aiming at showing that the proposed method can improve the robustness of the DLT-based FD approach. The FD results are shown in Figs. 9-12.

For the first experiment, in Fig. 9, at $T_d = 263$ rotor revolution, the faulty residual (purple line) becomes smaller than the normal residuals 1-3, thus the stall warning signal is obtained. However, the distance between the
Fault Diagnosis Based on Reinforcement Learning and Deterministic Learning

faulty residual and the normal residuals is small (such as the time $T_d < 260$ rotor revolution), which means that the algorithm can not totally distinguish the stall inception pattern from the normal pattern. In Fig. 10, the distance between the faulty and normal residuals is enlarged, which means that discriminative features have been extracted, since according to [29], the synchronization errors/residuals are the external reflections of the extracted dynamical features. More intuitively, in Figs. 7-8, it is seen that the features that the normal constant NN and fault constant NN learned are represented by constant RBFNN $\bar{W}_n^1(S(X))$ and $\bar{W}_f^1(S(X))$, respectively, and are mutually different. The learned features are spatially distributed due to the spatially localized learning property of the RBFNN.

For the second experiment, in Fig. 11, at $T_d = 28.4$ rotor revolution, the faulty residual (purple line) becomes smaller than the normal residuals 1-3, which leads to the false alarm. In Fig. 12, at $T_d = 278.6$ rotor revolution, the faulty residual (purple line) becomes smaller than the normal residuals 1-3, thus the rotating stall warning signal is obtained. Notice that the rotating stall occurs at approximately $T_d = 300$ rotor revolution, thus, a sufficient time (21.4 rotor revolutions) prior to the onset of rotating stall is provided based on the proposed FD approach.

5 Conclusion

In this paper, a novel RL and DLT based FD method has been proposed for a class of discrete-time nonlinear system with unknown system dynamics. The FD problem was addressed under the actor-critic structure, where the critic NN was utilized to approximate the strategic utility function. The action NN is updated using an DLT-based learning law to minimize the combined strategic utility function and the unknown system dynamics approximation error. According to the DLT, when the NN input trajectory is a period or recurrent orbit, the partial PE condition of the Gaussian RBFNN regression subvector is satisfied. Based on the partial PE condition, the exponential stability of the critic and action NN weights is proved using the Lyapunov stability theory. It
has been shown that this DLT-based learning mechanism not only considers
the approximation of the nonlinear system dynamics under the normal and
fault patterns, but also considers the extraction of discriminative features with
NN weight convergence. The extracted discriminative features improve the
robustness of the DLT-based FD approach. Experiment on the rotating stall
detection of the axial flow compressor shows the effectiveness of the proposed
approach.

Declarations

Some journals require declarations to be submitted in a standardised format.
Please check the Instructions for Authors of the journal to which you are
submitting to see if you need to complete this section. If yes, your manuscript
must contain the following sections under the heading ‘Declarations’:

- Funding: This work was supported in part by the Major Basic Program of
  Shandong Provincial Natural Science Foundation under Grant ZR2020ZD40,
  and in part by the National Natural Science Foundation of China under
  Grant 61890922.
Fault Diagnosis Based on Reinforcement Learning and Deterministic Learning

• Conflict of interest/Competing interests (check journal-specific guidelines for which heading to use): Not applicable
• Ethics approval: Not applicable
• Consent to participate
• Consent for publication
• Availability of data and materials: Not applicable
• Code availability: Not applicable
• Authors’ contributions: Not applicable

If any of the sections are not relevant to your manuscript, please include the heading and write ‘Not applicable’ for that section.

Appendix A  Proof of Theorem 1(i)

For simplification, the variables \( \Upsilon_i^0(K) \), \( \tilde{J}_i^0(K) \), \( \tilde{W}_{ic}^0(K) \) are denoted as \( \Upsilon(K) \), \( \tilde{J}(K) \), \( \tilde{W}_c(K) \) and \( \tilde{W}_a(K) \), respectively. The ideal NN approximation error is also neglected for brief.

Consider the following Lyapunov function

\[
V(K) = \alpha_c^{-1}\tilde{W}_c^T(K)\tilde{W}_c(K) + \frac{1}{\gamma_c}\|\xi(K-1)\|^2 + \Upsilon(K)^2 + \tilde{W}_a^T(K)\gamma^{-2}\tilde{W}_a(K) \tag{A1}
\]

where \( \xi(K-1) = \tilde{W}_c^T(K-1)S(K-1) \).

The difference of \( V(K) \) is

\[
\Delta V(K) = V(K+1) - V(K)
\]

\[
= \alpha_c^{-1}[\tilde{W}_c^T(K+1)\tilde{W}_c(K+1) - \tilde{W}_c^T(K)\tilde{W}_c(K)]
\]

\[
+ \gamma_c^{-1}(\|\xi(K)\|^2 - \|\xi(K-1)\|^2) + \Upsilon(K+1)^2 - \Upsilon(K)^2
\]

\[
+ \gamma^{-1}[\tilde{W}_a^T(K+1)\tilde{W}_a(K+1) - \tilde{W}_a^T(K)\tilde{W}_a(K)]
\]

\[
= \Delta V_1(K) + \Delta V_2(K) + \Delta V_3(K) \tag{A2}
\]

Define \( A = \tilde{W}_c^T(K)S(K), B = W_c^T S(K) + \alpha W_c^T S(K-1), C = \alpha \tilde{W}_c^T(K-1)S(K-1), D = \alpha^{N+1}P(K) \), then the first term in (A2) is expanded as

\[
\Delta V_1(K) = \alpha_c^{-1}[\tilde{W}_c^T(K+1)\tilde{W}_c^T(K+1) - \tilde{W}_c^T(K)\tilde{W}_c^T(K)]
\]

\[
\leq -2A(A + B + D - C) + \alpha_c\|S(K)\|^2(A + B + D - C)^2
\]

\[
= -(1 - \alpha_c\|S(K)\|^2)(A + B + D - C)^2 - A^2 + 3B^2 + 3C^2 + 3D^2
\]

\[
\leq -A^2 + 3B^2 + 3\alpha_c^2\|\xi(K-1)\|^2 + 3D^2 \tag{A3}
\]
Combining the first and second terms, we have

\[ \Delta V_1(K) + \Delta V_2(K) = -(1 - \frac{1}{\gamma_c}) \|\xi(K)\|^2 - (\frac{1}{\gamma_c} - 3\alpha^2) \|\xi(K - 1)\|^2 + 3B^2 + 3D^2 \quad (A4) \]

Choose the parameter \( \gamma_c \) such that \( 1 < \gamma < \frac{1}{3\alpha^2} \), then we have

\[ \Delta V_1(K) \leq -\lambda_1 \|\xi(K - 1)\|^2 + 3B^2 + 3D^2 \quad (A5) \]

where \( \lambda_1 > 0 \).

The third term in \((A2)\) is expanded along the difference equation \((31)\) as

\[ \Delta V_3(K) = \Upsilon(K + 1)^2 - \Upsilon(K)^2 
+ \gamma^{-1}[\bar{W}_a^T(K + 1)\bar{W}_a^T(K + 1) - \bar{W}_a^T(K)\bar{W}_a^T(K)] 
= (\alpha_t^2 - 1)\Upsilon^2(K) + 2\alpha_t^2 T_s \Upsilon(K)\bar{W}_a^T(K)S(K) + T_s^2(\bar{W}_a^T(K)S(K))^2 
- 2T_s\bar{W}_a^T(K)S(K)\Upsilon(K + 1) + T_s^2\gamma S^T(K)S(K)\Upsilon(K + 1)^2 \quad (A6) \]

Substituting \((31)\) into \((A6)\), we have

\[ \Delta V_3(K) = (\alpha_t^2 - 1 + T_s^2\gamma S^T(K)S(K)\alpha_t^2)\Upsilon^2(K) 
+ 2\alpha_t^2 T_s S^T(K)S(K)\bar{W}_a^T(K)S(K)\Upsilon(K) 
- T_s^2(1 - T_s^2\gamma S^T(K)S(K))(\bar{W}_a^T(K)S(K))^2 \quad (A7) \]

Since the assumption 3.3 holds, by using the well-known Young’s inequality, we have

\[ 2\alpha_t^2 T_s^3 S^T(K)S(K)\bar{W}_a^T(K)S(K)\Upsilon(K) 
\leq \alpha_t^2 T_s^3(\bar{W}_a^T(K)S(K))^2 + \|\Upsilon(K)\|^2 \quad (A8) \]

Then

\[ \Delta V_3(K) \leq -(1 - \alpha_t^2 - T_s^2\gamma s^2 \alpha_t^2 - \alpha_t^2 T_s^3 s^2)\Upsilon^2(K) 
- T_s^2(1 - \alpha_t^2 T_s s^2 - T_s^2\gamma s^2)(\bar{W}_a^T(K)S(K))^2 \quad (A9) \]

Choose the parameter \( \alpha_t \) and the learning gain \( \gamma \) such that

\[ 0 < \alpha_t < 1 \]
\[ \gamma < \min\{\frac{1 - \alpha_t^2}{T_s^2 s^2(\alpha_t^2 + \alpha T_s)}, \frac{1}{T_s s^2(\alpha_t + T_s)}\} \quad (A10) \]
Fault Diagnosis Based on Reinforcement Learning and Deterministic Learning

Then we have $1 - \alpha^2 - T_s^2 \gamma s^2 \alpha^2 - \alpha \gamma T_s^3 s^2 > 0$ and $1 - \alpha \gamma T_s s^2 - T_s^2 \gamma s^2 > 0$. Thus, $\Delta V_3(K)$ satisfies

$$\Delta V_3(K) \leq -\lambda_2 \Upsilon^2(K) - \lambda_3 (\tilde{W}_a^T(K) S(K))^2$$  \hspace{1cm} (A11)

where $\lambda_2, \lambda_3 > 0$.

Combing $\Delta V_1, \Delta V_2$ and $\Delta V_3$, we have

$$\Delta V(K) \leq -\lambda_1 \|\xi(K-1)\|^2 - \lambda_2 \Upsilon^2(K) - \lambda_3 (\tilde{W}_a^T(K) S(K))^2 + 3B^2 + 3D^2$$ \hspace{1cm} (A12)

This implies that $\Delta V(K) < 0$ as long as $\|\xi(K-1)\| > \sqrt{\frac{3(B^2+D^2)}{\lambda_1}}$, or $\|\Upsilon(K)\| > \sqrt{\frac{3(B^2+D^2)}{\lambda_2}}$, or $\|\tilde{W}_a^T(K) S(K)\| > \sqrt{\frac{3(B^2+D^2)}{\lambda_3}}$. This leads to the uniformly ultimate boundedness for $\xi(K-1)$, $\Upsilon(K)$, and $\tilde{W}_a^T(K) S(K)$. This ends the proof. \hfill \Box

References


Fault Diagnosis Based on Reinforcement Learning and Deterministic Learning


Fault Diagnosis Based on Reinforcement Learning and Deterministic Learning


