



Recent advances in data-driven dynamics and control

Zhi-Sai Ma^{1,2} · Xiang Li³ · Meng-Xin He^{1,2} · Su Jia^{1,2} · Qiang Yin^{1,2} · Qian Ding^{1,2}

Received: 1 July 2020 / Revised: 29 July 2020 / Accepted: 1 August 2020 / Published online: 8 August 2020
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

Abstract

With the rapid development of artificial intelligence, the data-driven methods have been extensively applied in the fields of machinery, power, civil engineering, transportation and other industries, such as fault diagnosis and prognosis. Whereas the data-driven based accurate dynamic modeling, system identification, and their applications in the design of engineering equipment and structures and health management for life-cycle process are still in the early stage. Currently, researches on nonlinear dynamics depend mainly on traditional methods of analysis. For the further advancement of the field of dynamics and control in the era of artificial intelligence, it is urgent to discover new knowledge representations and new predictions of complex dynamic laws from the observed data. The future research of dynamics and control must be driven by both the physical principles as well as data. This paper introduces some recent research work by our group in structural optimization, active vibration control, system identification, fault diagnosis and prognosis, and state identification of heart rate variability signal, by using data-driven methods.

Keywords Data-driven dynamics and control · Structural optimization · Active vibration control · System identification · Fault diagnosis and prognosis · Heart rate variability analysis

1 Introduction

Generally, data driven refers to a process or activity that is spurred on by data, as opposed to being driven by mere intuition or personal experience, or model-based analyses in science and technology community. With the advent of communication and sensor technologies, it is quite convenient to measure input–output data of engineering systems operated in complex environment. Because data is now easier to gather and inexpensive to store, big data analytics is gaining more ground as the best tool for decision making in the business, science and technology fields. This paper reports recent advances of data-driven engineering research in dynamics and control using various methods of machine learning.

Machine learning (ML), also known as statistical machine learning, is a branch of data science and artificial intelligence. Its basic idea is to build statistical model based on data, and use the model to analyze and predict data. Similar to artificial intelligence, ML is an interdisciplinary field, involving many basic disciplines including statistics, linear algebra and numerical calculation. ML can be divided into supervised learning, unsupervised learning and reinforcement learning. Supervised learning, including logistic regression, decision tree, support vector machines (SVM), K-Nearest Neighbor (KNN), naive Bayes, etc., uses labeled training data set for training, until the training convergence of the selected model. The training sample data of unsupervised learning does not have any label or output. Its purpose is to analyze the original data structure and find out the rules and relations between the data. Typical unsupervised learning tasks include clustering, dimensionality reduction, feature extraction, etc. Reinforcement learning emphasizes how to act based on the environment to maximize the expected benefits. Depending on continuous learning accumulation and evolution from the actual environment, reinforcement learning is also a technology that making machines obtain general intelligence.

✉ Qian Ding
qding@tju.edu.cn

¹ Department of Mechanics, Tianjin University,
Tianjin 300350, China

² Tianjin Key Laboratory of Nonlinear Dynamics and Control,
Tianjin 300350, China

³ College of Sciences, Northeastern University,
Shenyang 110819, China

Since 2006, deep learning (DL) has become a rapidly growing research direction of ML, redefining state-of-the-art performances in a wide range of areas such as object recognition, image segmentation, speech recognition and machine translation, as well as science and industrial technologies. Originated from the artificial neural network (ANN), DL is a branch of ML which is featured by multiple non-linear processing layers and tries to learn hierarchical representations of data. The trend of transitioning from traditional shallow ML methods to DL can be attributed to hardware evolution, algorithm evolution and data explosion [1]. DL approaches can also be categorized into the categories of supervised, semi- or partially supervised and unsupervised learning. Supervised learning approaches include Deep Neural Networks (DNN), Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN) including long short term memory (LSTM), etc. Semi-supervised learning is a kind of learning with partially labeled datasets. Deep reinforcement learning (DRL), Generative Adversarial Networks (GAN), RNN including LSTM and Gated Recurrent Units (GRU) are used as semi-supervised learning techniques. Un-supervised learning approaches include Auto Encoders (AE), Restricted Boltzmann Machines (RBM), and the recently developed GAN. A key difference between traditional ML and DL is in how features are extracted [2]. Traditional ML approaches use handmade features by applying several feature extraction algorithms including Scale Invariant Feature Transform (SIFT), Speeded Up Robust Features (SURF), Empirical mode decomposition (EMD), etc. The learning algorithms including SVM, Random Forest (RF), Principle Component Analysis (PCA) and many more are applied for classification on the extracted features. For DL, the features are learned automatically and are represented hierarchically in multiple levels. The advantage of DL over traditional ML is increased especially with the dataset dimension [3].

ML and DL are employed in situations where machine intelligence would be useful for cases: (1) absence of a human expert, or when human are unable to explain their expertise; (2) solution to the problem changes over time; (3) solutions need to be adapted to particular cases; and (4) the problem size is too large for our limited reasoning capabilities. A number of complex case studies, chosen from diverse scientific disciplines, are presented to illustrate the power of DL methods for enabling scientific analytics tasks and discovery, including galaxy shape modeling with probabilistic auto-encoders, extreme weather events in climate simulations, learning patterns in cosmology mass maps, decoding speech from human neural recordings, clustering Neutrino experiment data of Daya Bay Reactor with denoising auto-encoders, new physics events classifying at the Large Hadron Collider. Open challenges in practical applications of DL are also discussed, such as performance and scaling (dealing with expensive computation), complex data, lack of labeled

data, hyperparameter tuning (it is important to adopt automatic tuning techniques), and interpretability [4].

Differential equations, including ordinary differential equations (ODEs) and partial differential equations (PDEs), are used to describe the dynamical nature of moving objects, structures and even social phenomena. Besides analytical and numerical methods, ANNs can also be adopted to solve ODEs and PDEs [5]. It is one of the research hotspots in recent years to replace the continuous solution of differential equations with DNN mappings, in which the advantage of the nonlinear representation capabilities of DNNs is well used. That candidate solution of the equation is a direct output of neural networks. The distance between the output of neural networks and the true solution can be minimized by adjusting the coefficients of the neural networks. Chen [6] proposed a DNN model to solve ODEs. Conventional algorithms are used to verify the output of the network, and the network model is finally trained as an end-to-end differential solver. Wei [7] proposed a rule-based self-learning approach using DRL to solve nonlinear ODEs and PDEs. The solver consists of a DNN-structured actor that outputs candidate solutions and a critic derived only from physical rules (governing equations, boundary and initial conditions). Solutions in discretized time are treated as multiple tasks sharing the same governing equation, and the current step parameters provide an ideal initialization for the next, which shows a transfer learning characteristic. The Navier–Stokes, van der Pol and Lorenz equations are solved to show the high accuracy and efficiency of the solution process.

For the researches of dynamics and control in science and technology, we need to deal with complex systems in complex environment. The complexity includes: (1) nonlinearity resulted from complex evolution of damage and degradation of structures and materials; (2) uncertainty of operation condition and environment; (3) coupling between loads and structural nonlinear characteristics; and (4) limited cognition for mixed information. For instance, common joints between substructures often feature very complex physics, including heterogeneous stick–slip behavior at the microscopic level, hysteresis, Hertzian contact and local concentrations of stresses and strains. This makes it almost impossible to specify an accurate and physically-motivated model in terms of macroscopic nonlinear stiffness and damping lumped elements. Black-box models, which incorporate no prior knowledge but take advantage of a sufficiently rich and flexible mathematical structure to capture all relevant dynamics information in measured data, are useful in these situations. Over the past decade, NN-based identification has remained the most popular black-box modeling technique within the structural dynamics community [8]. However, most researches on dynamics and control are mainly based on ideal hypothesis and principle analysis up to now. Dynamic experimental data accumulation, information deep mining and data-driven based investigations on dynamics

of important industrial equipment are still in the initial stage, which cannot meet the requirements of complex system identification, modeling accuracy, dynamical design and health management.

In modern industries, machinery fault diagnosis and prognosis have been of great importance in many practical applications, such as manufacturing, transportation, aerospace industry etc. Significant economic benefits and enhanced operating safety are expected if accurate and reliable fault diagnosis methods are available. Fault diagnosis and prognosis are based on the fault mechanism analysis, which is to get the relationship between the fault signal of equipment and the parameters of equipment system through theoretical or a large number of experimental analyses. For the mechanical dynamic system, the most commonly used state information is vibration signal. So ML and DL based fault diagnosis technology can be also regarded as the data-driven dynamic application.

Traditionally, the physical model-based signal processing methods have been popularly and successfully used in machinery fault diagnosis and prognosis problems. Since the machines are becoming more and more complex in the recent years, it is difficult to build accurate physical models for diagnostics, and the conventional approaches have thus been less effective in practice. Conventional ML methods, including ANNs, PCA, SVM, etc., have been successfully applied to detecting and categorizing bearing faults, while the application of DL methods has sparked great interest in both the industry and academia in the last 5 years. The DL-based methods over the conventional ML methods are analyzed in [1] in terms of metrics directly related to fault feature extraction and classifier performances. Applications of the most popular DL approaches, such as AE, RBM, CNN and RNN, in machine health monitoring systems are reviewed in [9]. Some new trends of DL-based machine health monitoring methods are also discussed.

In this survey paper, our group's recent research work based on data-driven methods is presented to illustrate the innovations in the field of dynamics and control. The work includes structural optimization, active vibration control, system identification, fault diagnosis and prognosis, and state identification of heart rate variability signal. Finally, future directions for the related topics are discussed.

2 Multi-objective optimization

Generally, a multi-objective optimization problem (MOP) can be defined as [10],

$$\min_{\mathbf{x} \in Q} \{\mathbf{F}(\mathbf{x})\} = \min_{\mathbf{x} \in Q} \{f_1(\mathbf{x}), \dots, f_n(\mathbf{x})\}, \quad (2.1)$$

where $\mathbf{x} \in Q \subset \mathbf{R}^m$ is a design vector with m dimensions. $\mathbf{F}(\mathbf{x})$ consists of n objective functions, namely $\mathbf{F} : \mathbf{R}^m \rightarrow \mathbf{R}^n$.

Different from the single-objective optimization, there is no single solution that can minimize all the functions simultaneously due to the conflicting objectives. Therefore, the concept of Pareto dominance is widely used in the field of MOPs. The solutions of the MOP are called the Pareto set \mathcal{P} and the image $\mathbf{F}(\mathcal{P})$ or \mathcal{PF} is called the Pareto front. The solutions obtained by the MOP algorithms are also termed non-dominated. The Hypervolume performance measure $I_H(\mathcal{PF}^*)$ [11] can be used to express the quality of the non-dominated solutions \mathcal{PF}^* by the MOP algorithm. A high value of I_H means that the approximate non-dominated solutions \mathcal{PF}^* are close to the real Pareto front \mathcal{PF} . More details of the MOP can be found in [10].

There are many methods for solving MOPs such as NSGA-II [12], MOPSO [13], and cell mapping method [10]. In dealing with structural optimization problems, many objective functions involve computationally expensive simulations, and the optimization methods often require many iterations. Generally, engineering optimization to meet multiple and possibly conflicting objectives leads to a MOP [14, 15], because the constraints, larger populations [16] and more generations [17] are needed to obtain more accurate solutions. However, the computational cost of the objectives in some structural optimization problems is prohibitively high, such as fluid dynamic simulations and the high-fidelity finite element analysis. Therefore, it is challenging for optimization methods, such as the genetic algorithm [16, 17] and the cell mapping method [10, 18, 19], to cope with large scale engineering optimization problems due to the high computational cost of the objective functions. In order to reduce the computational cost, data-driven methods have been widely used to improve the computational efficiency [20].

In this section, we present the application of ML method in structural optimization. A generation-based support vector regression (SVR) assisted framework for multi-objective optimization is first proposed. Then, two data-driven optimization problems are presented as demonstrations to show the efficiency of the data-driven optimization method, including functionally graded beam optimization and design of periodic beam structures.

2.1 Generation-based SVR-assisted framework

In data-driven optimization problems, surrogate models, including support vector regression [21], neural networks [22] and radial basis function networks [23] are trained by using the database evaluated by the real objective functions with a limited number of evaluations. Then, the MOP algorithms, such as NSGA-II [12] and MOEA/D [24], are used to obtain the non-dominated solution by using the surrogate models. In this paper, we employ support vector regression [21] as surrogate models.

A generation-based SVR-assisted framework for multi-objective optimization is proposed. The goal is to obtain an approximation non-dominant solution set A to the real Pareto solution. The main steps are given, as follows,

- (1) Create the initial design using Latin hypercube sampling, calculate the exact fitness values for every point and generate the initial database $D_{global} = \{(x_1, y_1), \dots, (x_{N_t}, y_{N_t})\}$.
- (2) Use the database to train the surrogate models based on the SVR.
- (3) Obtain the approximate Pareto front $P f^*$ and Pareto set $P s^*$ by using the well-known NSGA-II [12].
- (4) Select potential points from $P f^*$ and $P s^*$ by using the concept of Pareto dominance, and calculate the responses by using the real fitness function. All the data evaluated by the real fitness function are stored in D_{all} .
- (5) Update the non-dominant solution set A and the database D_{global} by using the selected points.
- (6) Determine whether to stop searching by monitoring the value of I_H and the number of fitness function evaluations. If the stopping criterion is met, output the non-dominant solution set A, otherwise go to step (2).

Note that there are two options of the database for training the surrogate models in step (2), namely D_{global} or D_{local} . The database D_{global} has a limitation of the capacity N_t . We select the solutions from the non-dominant solution set A to update D_{global} at each iteration. Because the archive A is non-dominant solutions, the surrogate models trained by using D_{global} has good generalization ability. During the initial iteration D_{global} is used to speed up the global search process. However, when falling into the local optimum, the surrogate models maybe not a suitable choice to predict the response of the local direction. Therefore, we use the weighted sum approach [24] to convert the MOP to a single optimization problem (SOP). After that, we select the best N_t solutions to the SOP from D_{all} by using Eq. (2.2) as the local database D_{local} .

$$\text{ming}(x) = \sum_{i=1}^n \lambda_i f_i(x) \tag{2.2}$$

where λ_i is a coefficient vector and will be randomly generated. The surrogate models by using D_{local} will be more precise than those by D_{global} in the local of the design space. If the performance metric I_H changes slowly, the local search will be performed in the next iteration for escaping the local optimum.

To obtain a more accurate surrogate model, a cooperative way is adopted to combine the three SVR models by using different kernels

$$f = \omega_1 \cdot f_G + \omega_2 \cdot f_{MQ} + \omega_3 \cdot f_{Linear} \tag{2.3}$$

where f_G , f_{MQ} and f_{Linear} are the SVR models using the gaussian, the multi-quadratic and the linear kernels, respectively. ω_1 , ω_2 and ω_3 are the weights of the three models, given by

$$\omega_i = \frac{\sum_{i=1}^3 MSE_i - MSE_i}{\sum_{i=1}^3 MSE_i}, \quad i = 1, 2, 3 \tag{2.4}$$

where MSE_1 , MSE_2 and MSE_3 are the mean square error of the three SVR models in the training process.

2.2 Optimization of FGM beams

Optimal design of beam structures for better acoustic performance is an interesting research subject [25]. Changing material or geometric parameters can be an effective approach for noise reduction. Functionally graded materials (FGM) have spatially continuous variations of properties, which can be tailored to improve the ability of structures to reduce sound radiation [26].

Suppose that the FGM beams are made of aluminum and steel. The volume fraction of material aluminum is an arbitrary and continuous function of the spatial coordinate x . A number of the sampled volume fraction $V1(x_i)$ of material aluminum along the beam are taken as the design variables, i.e. $V_{1i} = V_1(x_i)$. Then the smooth function $V1(x)$ is constructed by using the spline interpolation. Note that the spline function in this section is C1 continuous. Ten sampled points are taken as design variables such that the dimension of the design space Q is 10, as follows

$$Q := \{k \in \mathbb{R}^{10} | 0 \leq k_i \leq 1, i = 1, 2, \dots, 10\} \tag{2.5}$$

The aim is to find the optimal material distribution that meets two conflicting objectives defined as

$$f_1 = \frac{1}{\omega_u - \omega_l} \sum_{i=1}^M \Delta \omega P_{dB}(\omega_i, x), \quad f_2 = Mass(x) \tag{2.6}$$

where P_{dB} represent the radiated sound power level (SPL) in dB. The first objective is defined to minimize the averaged SPL over a range of frequencies, where ω_u and ω_l are the lower and upper bounds of the frequencies of interest respectively. The second objective is the weight of the structure. The detail of dynamics modeling can be found in [26].

We consider the optimization problem of a clamped-clamped FG beam, made of aluminum and steel, being excited by a harmonic load with an amplitude of 5 N at point $x = L/\sqrt{2}$. The length, height and width of the beam are 1 m, 0.01 m and 0.05 m respectively. Setting $\omega_u = 155$ Hz and $\omega_l = 145$ Hz, the first objective function of this

example is defined to reduce the SPL of the frequency in the range of [145, 155] Hz.

Figure 1 shows the Pareto front obtained by the proposed method, consisting of 275 solutions. The computational budget is 3064 fitness function evaluations. In the subfigures, points 1–3 are examples of the optimal designs demonstrating the volume fractions of aluminum of the three designs. It can be seen that the density and Young’s modulus are increased at the location of the load. By adjusting the local material properties of the FGM beam, the structural–acoustic behavior is redesigned according to the optimization objectives.

Figure 2 illustrates the radiated sound power of the aluminum beam, the steel beam and one optimal solution in the Pareto front in Fig. 1 (point 3). In the frequency range of interest, the average SPL of the aluminum beam, the steel

beam and the optimal beam are 88.51, 78.95 and 67.17 (dB), respectively.

2.3 Optimization of beams with embedded ABH structure

Recently, acoustic black hole (ABH) structure has been widely studied in the field of vibration control, energy harvesting and noise reduction [27, 28]. The ABH effect is based on the peculiarity that the velocity of bending waves propagating along the tapered beam reduces gradually and approaches zero theoretically at the zero thickness tip [29, 30]. Therefore, the geometric parameters are critical for the dynamic behaviors of the ABH structures [31].

In this case, the vibration and sound radiation are analyzed by using the spectral element method. A data-driven optimization study of the ABH parameters to maximize the sound transmission loss (STL) is presented. Suppose the amplitude of the acoustic incident wave is 1 Pa and the incidence angle $\pi/2$ with respect to the neutral axis of the beam. The incident sound power W_i over the bottom of the beam and the transmitted sound power W_t are given by

$$W_i = \frac{L}{2\rho_{air}c_{air}}, W_t = \frac{1}{2\rho_{air}c_{air}} \int_0^L |p_t^2(x)| dx \tag{2.7}$$

where ρ_{air} and c_{air} are the air density and sound speed in the air, L the length of the periodic beam, p_t the transmitted sound pressure in the upper face of the beam. Accordingly, the STL is given by

$$STL = 10 \log_{10} \frac{W_i}{W_t} \text{ (dB)} \tag{2.8}$$

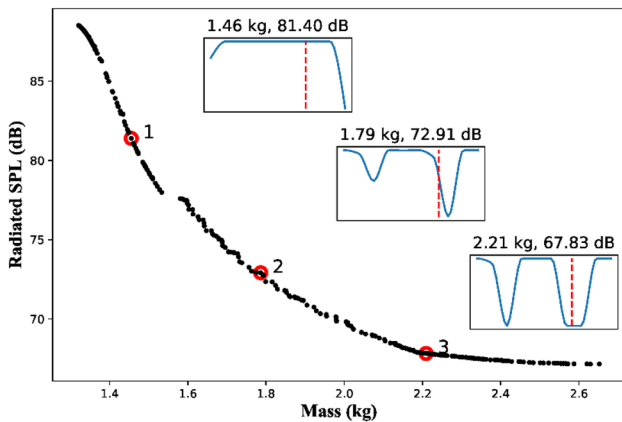
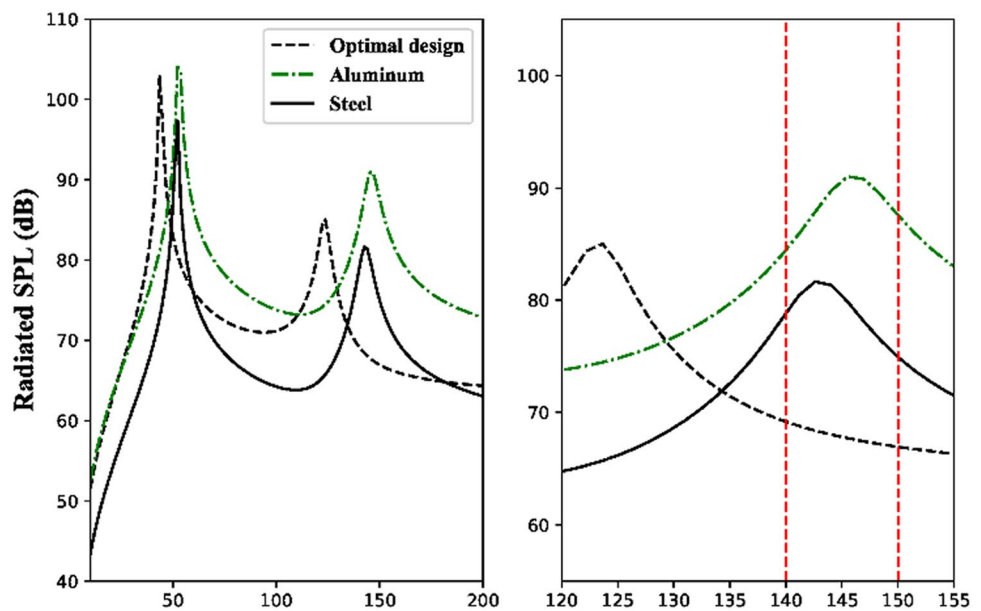


Fig. 1 The Pareto front for a clamped–clamped beam under a point load (the red dashed line)

Fig. 2 The radiated SPL of the aluminum, the steel and the optimal beams



Consider a clamped–clamped beam with five identical ABH units. Two objectives are defined as

$$\begin{cases} f_1 = Mass(x) \\ f_2 = \frac{1}{\omega_{ub}-\omega_{lb}} \int_{\omega_{lb}}^{\omega_{ub}} -STL(\omega, x) d\omega \end{cases} \quad (2.9)$$

where f_2 means that minimizing the opposite of the STL of the structure. In this case, we set $\omega_{ub} = 150$ Hz and $\omega_{lb} = 700$ Hz.

Figure 3 shows the design variables $X = [x_1 x_2 x_3 x_4 x_5 x_6 x_7]$ of a single ABH unit. x_1 and x_2 are the length of the unit and the ABH region, respectively. x_3 and x_4 represents the thickness of the uniform region and the thinnest part of the ABH region, respectively. x_5 is the value of exponent m . The length and the thickness of the damping layer are represented by x_6 and x_7 . The design space Q is given by

$$Q := \{k \in \mathbb{R}^7 | lb_i \leq x_i \leq ub_i, i = 1, 2, \dots, 7\} \quad (2.10)$$

where $lb = [0.08, 0.04, 0.005, 0.0004, 2, 0, 0.0001]$ (m) and $ub = [0.12, 0.08, 0.012, 0.002, 5, 0.05, 0.0004]$ (m).

Figure 4 shows the Pareto front obtained by the proposed method, consisting of 231 solutions. In the subfigures, points 1–3 are examples of the optimal designs demonstrating the single ABH unit. The gray and the red areas represent the beam structure and the damping layer, respectively. It can be seen that the Pareto front provides a wide range of design options. For example, point 2 is very similar to the structure in [31].

Figure 5 shows the STL of the ABH beam and the uniform beam. The solid line and dashed line represent the optimal ABH beam (point 2 in Fig. 4) and the uniform beam with the same mass, respectively. Recall that the objective is to maximize the STL in the frequency range [150, 700] Hz. We can see that the STL of the ABH beam is improved in the frequency range [150, 700] Hz. Note that the first “band gaps” of the beams with ABH in [31] is 150–850 Hz.

Figure 6 shows the sound pressure levels distribution of the ABH beam and the uniform beam at 533 Hz, which is within the band gap of the ABH beam and the 3rd mode of

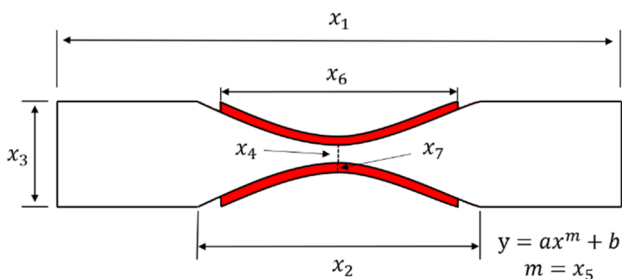


Fig. 3 A single ABH unit

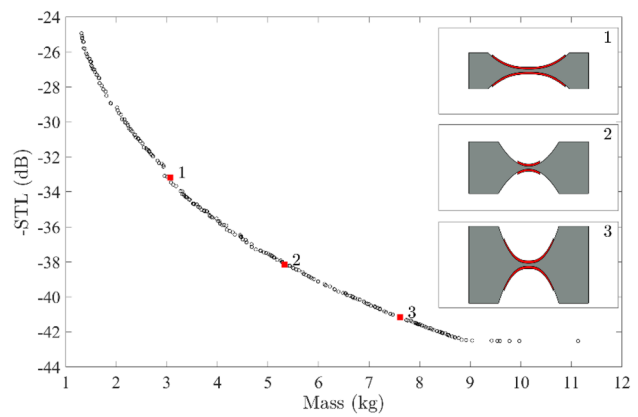


Fig. 4 The Pareto front obtained by the proposed method, consisting of 231 solutions

the uniform beam, respectively. Evidently, the sound pressure levels are significantly reduced due to the ABH effect.

3 ADP-based active vibration control

With the rapid development of information science, sensor technology, and artificial intelligence, lots of new data-driven control methods have emerged. Different from the traditional control methods whose control effect is heavily influenced by the model precision, the data-driven control design methods are free from the constraints of the system model and control effect depend only on the online or offline state data. The optimal control problem needs to solve the Riccati equation of the linear system or the Hamilton–Jacobi–Bellman (HJB) equation of the nonlinear system, both contain system model information [32]. Dynamic programming is a classical technique to solve the nonlinear HJB

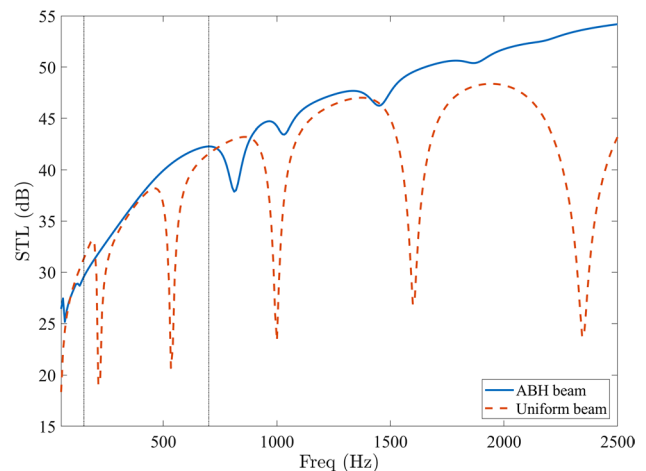
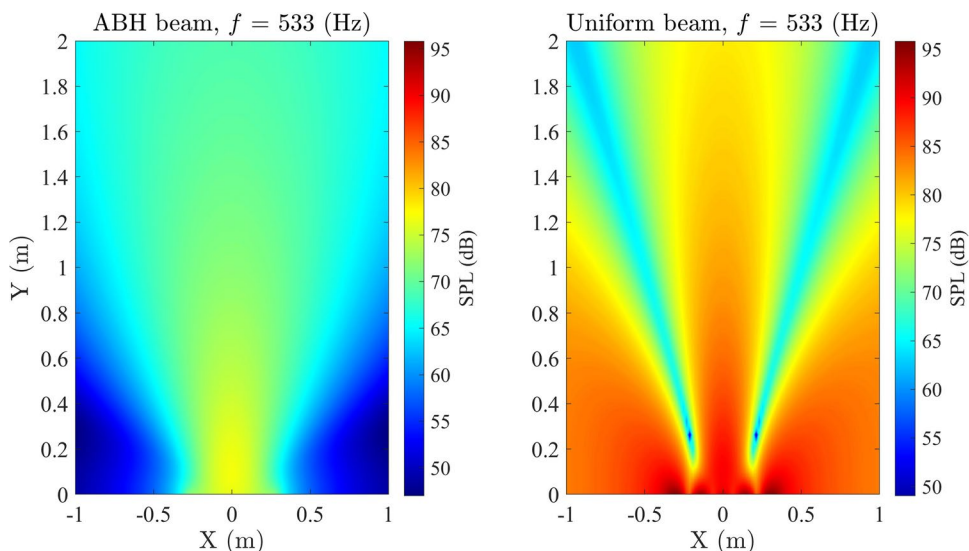


Fig. 5 STL of the ABH beam and the uniform beam

Fig. 6 Sound pressure levels of the ABH beam and the uniform beam at 533 Hz



equation, which often leads to dimension curse problem for its backward solution characteristics. Adaptive dynamic programming (ADP) combines the thought of reinforcement learning and takes advantage of neural networks to approximate function [33]. The actor-critic network is a typical architecture used to solve the optimal control without depending on the system model. Vrabie and Lewis proved the stability and convergence of the online ADP algorithm [34]. As an effective data-driven controller design method, ADP algorithms have been applied in many fields, such as power system [35] and aircraft control [36].

The data-driven ADP method solves the HJB equation through policy iteration algorithm based on input and output data instead of depending on the knowledge of system dynamics. And the typical actor-critic structure is used to approximate the cost function and control policy. In this section, the data-driven ADP-based control method is presented and applied to an active mass damping (AMD) system and a nonlinear two-dimensional airfoil system, to illustrate its good vibration control effect on the base-excited and self-excited structural vibrations with unknown model parameters.

3.1 Data-driven policy iteration algorithm

For a nonlinear dynamical system

$$\dot{x} = f(x) + g(x)u \tag{3.1}$$

where $x(t) \in R^n$ is state variable vector, $u(t) \in R$ is control variable. The system functions $f(\cdot) \in R^n$ and $g(\cdot) \in R^{n \times 1}$ are assumed to be unknown matrices and are differentiable in the arguments satisfying $f(0) = 0$. Based on optimal control theory, we define the infinite horizon cost function as

$$V(x(t_0)) = \int_{t_0}^{\infty} (x^T Q x + u^T R u) dt \tag{3.2}$$

where Q and R are symmetric positive definite matrix.

The Hamiltonian function of the problem is

$$H(x, u, V) = r(x(t), u(x(t))) + (\nabla V_x)^T [f(x) + g(x)u] \tag{3.3}$$

where $r(x(t), u(x(t))) = x^T Q x + u^T R u$ and ∇V_x denotes the gradient of the cost function.

The optimal feedback control law can be determined by

$$u^*(x(t)) = -\frac{1}{2} R^{-1} g^T(x) \nabla V_x^* \tag{3.4}$$

Combining the Hamiltonian function and optimal control law, the formulation of the HJB equation is obtained as

$$(\nabla V_x^*)^T f(x) + Q(x) - \frac{1}{4} (\nabla V_x^*)^T g(x) R^{-1} g^T(x) \nabla V_x^* = 0 \tag{3.5}$$

However, solving this equation is generally difficult and it also always depend on the complete knowledge of the system dynamics. The data-driven policy iteration is applied to solve the HJB equation by using the input and output data of the system to learn the solution of the HJB equation. Rewrite the system as

$$\dot{x} = f + g u^{(i)} + g [u - u^{(i)}] \tag{3.6}$$

where $u^{(i)}$ is the control policy of each iteration and integrate both sides of cost function among the interval $[t, t + \Delta t]$, we have

$$\begin{aligned} & V^{(i+1)}(x(t)) - V^{(i+1)}(x(t + \Delta t)) \\ &= -2 \int_t^{t+\Delta t} [u^{(i+1)}(x(\tau))]^T R [u^{(i)}(x(\tau)) - u(\tau)] d\tau + \int_t^{t+\Delta t} r(x, u^{(i)}) d\tau \end{aligned} \tag{3.7}$$

where $V^{(i+1)}(x)$ and $u^{(i+1)}(x)$ are the unknown function to be determined. In this equation there is no system model information. In fact, the system dynamics are embedded in the measurement of the state and control signal. Given an initial admissible control policy, the control strategy sequence can be finally obtained by using the policy evaluation step and the policy improvement step alternately.

3.2 Actor-critic neural network structure

The actor-critic neural network structure is introduced to implement the data-driven ADP method. The cost function and control policy can be approximated by this structure. The control policy and cost function are replaced by two separate neural network, actor neural network and critic neural network [37]. Let $\phi(x)$ and $\psi(x)$ be the vector of linearly independent activation functions for critic NN and actor NN, with L_v and L_u being the number of hide layer neurons. Then the cost function and control law are given by

$$\hat{V}^{(i)}(x) = \sum_{j=1}^{L_v} \theta_{V,j}^{(i)} \phi_j(x) = \phi^T(x) \theta_V^{(i)} \quad (3.8)$$

$$\hat{u}^{(i)}(x) = \sum_{k=1}^{L_u} \theta_{u,k}^{(i)} \psi_k(x) = \psi^T(x) \theta_u^{(i)} \quad (3.9)$$

where θ_V and θ_u are weight vectors of critic and actor NNs respectively. After simplification and formula transformation, the residual error can be obtained as

$$\sigma^{(i)}(x(t), u(t), x(t + \Delta t)) = \bar{\rho}^{(i)}(x(t), u(t), x(t + \Delta t)) \theta^{(i+1)} - \pi^{(i)}(x(t)) \quad (3.10)$$

where $\bar{\rho}^{(i)}$ and $\pi^{(i)}$ contain system data information and no system model.

The weights $\theta^{(i+1)}$ can be updated by using least squares, gradient descent or weighted residual method [38]. Finally, the unknown critic and actor NNs weight vector $\theta^{(i+1)}$ can be determined. After achieving the convergence of the weights sequence, the approximate optimal control policy can be obtained. It should be noted that the data-driven ADP method is able to obtain the approximated optimal control strategy based on system input and output data through the policy iteration algorithm and actor-critic neural network structure, even if the system is subject to external disturbances.

3.3 Case studies

The data-driven ADP-based control method is applied to two typical examples including an AMD system and a nonlinear two-dimensional airfoil system. Figure 7 shows a AMD active control system with mass of the structure

M_f , mass of the controlling car M_c , displacement of the structure relative to ground x_f , displacement of the car relative to the structure x_c , and the seismic acceleration \ddot{x} , respectively. Define the system state as $x = [x_c x_f \dot{x}_c \dot{x}_f]^T$, the system model can be written in the state space form

$$\dot{x}(t) = Ax(t) + Bu(t) + D_g \ddot{x}_g(t) \quad (3.11)$$

In the control simulation, the critic and actor NNs should be constructed properly to fit different practical problems. After the data collection, offline iteration is used to learn the optimal weight vector. The vibration control simulation of AMD was carried out for Kobe seismic wave. The peak acceleration of the Kobe wave is 0.14 g and the duration is 63.7 s. The simulation results are shown in Fig. 8. It shows that the maximum displacement and acceleration are reduced from 12.8255 to 5.0592 mm and 3.6186–1.7517 m/s², with reduction of 57% and 51.6% respectively. The active control strategy obtained by the robust data-driven ADP method effectively reduces the vibration response of the slender structure under an earthquake.

The second examples is a typical two-dimensional airfoil with plunge displacement h and the pitch angle α , as shown in Fig. 9. Stability analysis indicates that the flutter happens as the critical value 10.23 m/s is exceeded by flutter velocity U . Applying the data-driven ADP-based control method to the system for controlling the airfoil flutter, the simulation results are provided in Fig. 10, which shows that the trajectories of pitch angle and plunge motions under the action of the controller converge to zero in a few seconds.

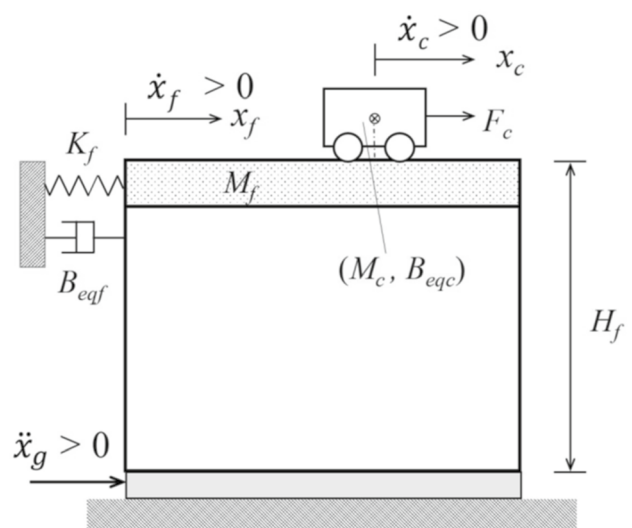


Fig. 7 AMD active control model of structural vibration

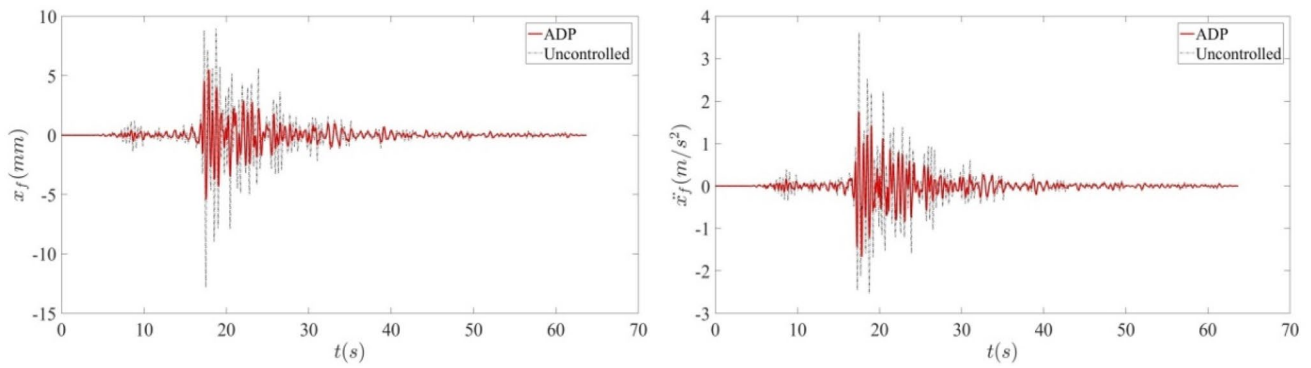


Fig. 8 Time histories of the structure displacement and acceleration under earthquake

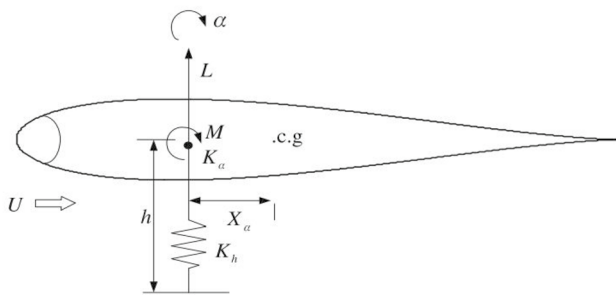


Fig. 9 A dynamical model of the two-dimensional typical airfoil

4 System identification

With the development of engineering applications, more and more complex structures are being used and their intrinsic time-varying/nonlinear characteristics are increasingly inevitable, which brings many problems on structural dynamics [8, 39–42]. It is usually difficult to build an explicit model of a complex structure by exclusively using mechanism analysis, and there is also no guarantee that the model will accurately represent its

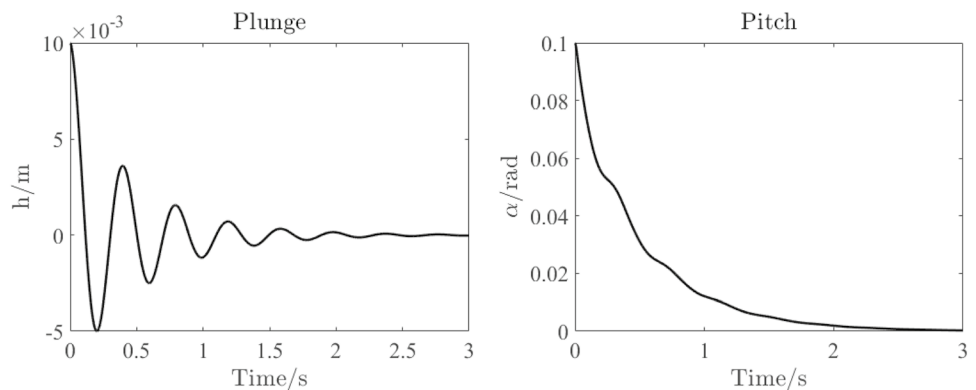
time-varying/nonlinear dynamic characteristics. As an inverse problem, system identification is worth more attention and investigation to obtain the dynamic characteristics of a structure under operating conditions based on its measured data, which is also referred to as data-driven system modeling.

In the past three decades, the system involved in the identification domain has been expanding to time-varying/nonlinear systems, which presents a great challenge to identification researches. In this section identification of time-varying/nonlinear systems and hybrid modeling of nonlinear jointed structures are respectively addressed.

4.1 Time-varying system identification

To pursue more accurate modeling and analysis of time-varying systems, the output-only recursive identification on the measured non-stationary signals via parametric time-domain methods is mainly adopted due to the following reasons [42–45]: (1) able to determine the dynamics of system by only using the measured responses but without difficult-measured excitation forces; (2) much more efficient than batch identification by allowing the real-time signal processing to keep up with the data acquisition; (3) representation

Fig. 10 The response of airfoil with the controller at $U = 16$ m/s



parsimony, achievable accuracy, resolution and tracking of the time-varying dynamics as the parametric identification method; and (4) able to avoid the truncation errors caused by domain transformation as the time-domain identification method.

Recently, ML methods have been used for time-varying system identification. For example, the idea of kernel recursive extended least squares was adopted to estimate the time-dependent autoregressive moving average (TARMA) model parameters, which achieves superior estimation accuracy, lower computational complexity and enhanced tracking capability than the existing recursive pseudo-linear regression method [43]. Besides, Bayesian linear regression [46, 47], ridge regression [44, 48], and least squares support vector machine [49] have also been utilized and extended to develop novel estimators for time-varying system identification, so as to improve achievable accuracy, reduce computational complexity and/or meet some specific identification requirements.

Two examples with typical time-varying dynamics and their corresponding identification results by multivariate recursive Bayesian linear regression are here illustrated [47]. As shown in Fig. 11, a simply supported beam with moving mass can be viewed as a simplified model of the vehicle-bridge interaction system. The estimated natural frequencies from the 30 tests and the estimated auto power spectral density (PSD) of the acceleration at No.4 measuring point from the single test are presented in Fig. 12, where baseline natural frequencies are denoted by the red dashed lines. The fluid–structure interaction systems with variable mass distribution widely exist in the field of aerospace applications. A liquid-filled cylindrical structure with decreasing filling mass can be used to model and analyze the slender aerospace structure with decreasing fuel, as shown in Fig. 13. Similarly, the estimated natural frequencies from the 30 tests and the estimated auto PSD of the acceleration at No.5 measuring point from the single test are depicted in Fig. 14. Evidently, natural frequencies are adequately

estimated with good time-varying tracking accuracy, and the estimated PSD ridges are in good agreement with the baseline natural frequencies.

4.2 Nonlinear system parameter identification

Nonlinear systems exist widely in the real world and engineering field. For example, freeplay nonlinearity is inevitable for deployable structures due to the factors such as mis-machining tolerance, assembly error, and abrasion. Friction nonlinearity caused by contact and sliding between bodies exists widely in mechanical systems. Kerschen et al. [39] and Noël et al. [8] reviewed the research progress of nonlinear system identification in structural dynamics over the past decades and presented a critical survey of parameter estimation methods including linearization, time- or frequency-domain methods, time–frequency analysis, nonlinear modal analysis, black-box modeling, and model updating methods.

Multi-stable solution is an important difference between nonlinear systems and linear systems. In order to solve this kind of nonlinear parameter identification problem, a feature-based parameter identification framework is established. As shown in Fig. 15, two steps are proposed to identify the multi-stable solutions in parameter estimation by data-driven methods. Step 1, one classifier is used to classify multiple solutions and judge the corresponding solution set of response data. Step 2, the corresponding model of solution set is selected to identify the nonlinear parameters. In time domain, there are many statistical features covering a wide range of popular time domain characteristics to be extracted from the preprocessed vibration signals, such as mean value, average absolute value, root-mean-square value, variance value, skewness coefficient, kurtosis coefficient, etc. These statistical parameters can be used to construct the identification model. The sliding time window approach [50] is widely adopted for data segmentation to ensure that the features extracted in the time window are relatively stable for the system to change over time and reduce the operational

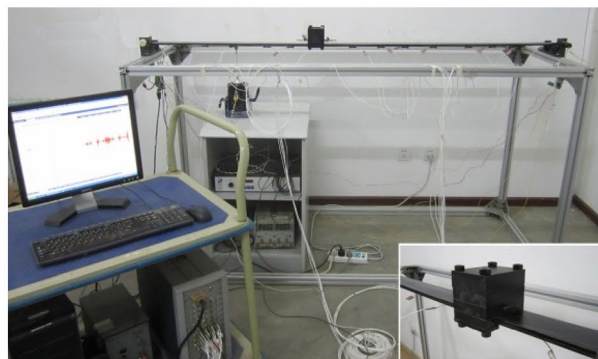
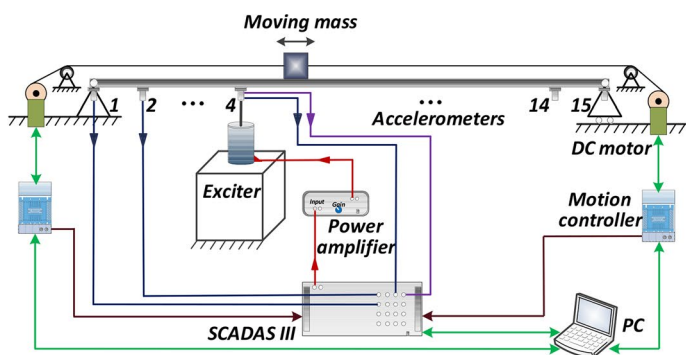


Fig. 11 A simply supported beam carrying single moving mass

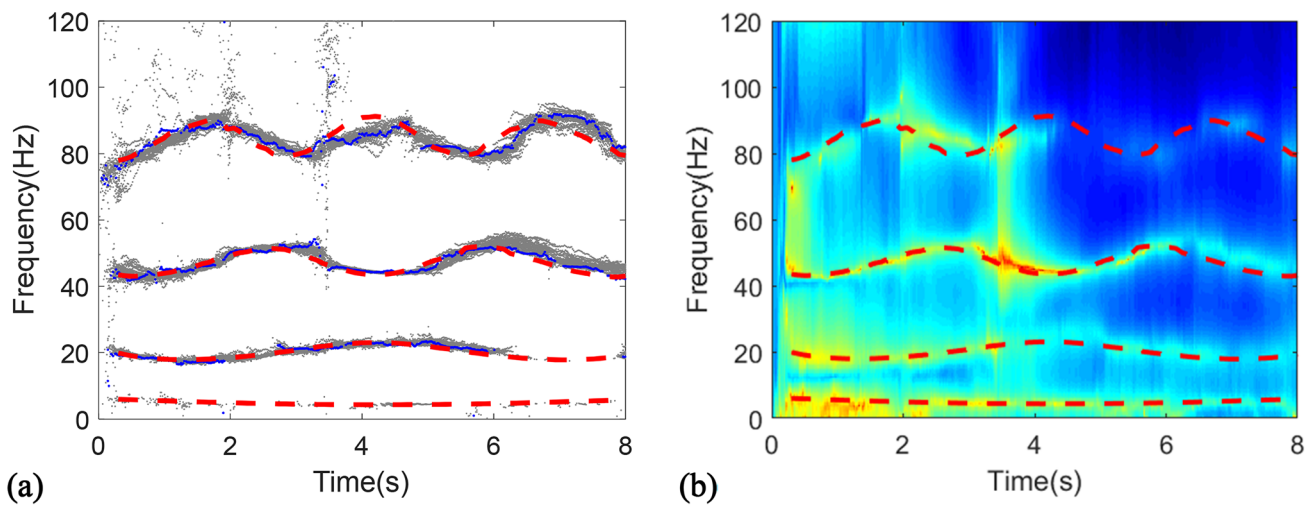
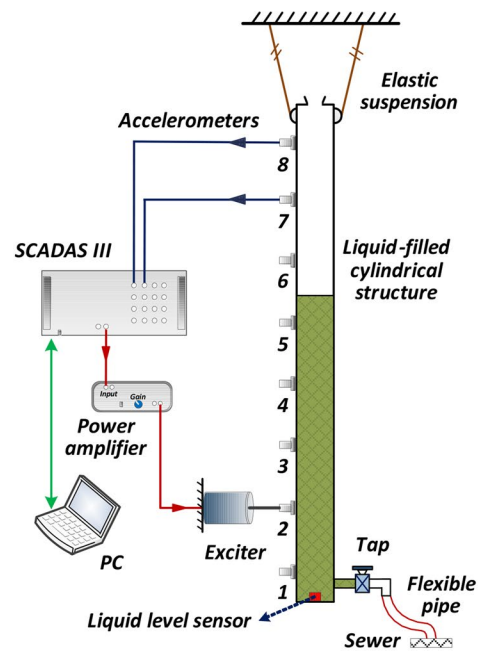


Fig. 12 Natural frequency and PSD estimates (gray dots: estimated natural frequencies from the 30 Monte Carlo tests; red dashed lines: baseline natural frequencies). (Color figure online)

Fig. 13 A liquid-filled cylindrical structure with decreasing filling mass



difficulty of feature extraction. Generally, the selection of window size is related to the period of solution, and different window sizes will affect the results of feature extraction.

Considering a Duffing oscillator governed by

$$m\ddot{x}(t) + c\dot{x}(t) + kx(t) + dx^3(t) = p(t) \quad (4.1)$$

where the mass $m = 1$ kg, the damping $c = 2$ Ns/m, the stiffness coefficient $k = (10\pi)^2$ N/m, and the external excitation $p(t) = A\cos(2\pi ft)$, where $A = 500$ N and f is increased from 0 until the resonance is passed through. The response of

nonlinear system depends on the initial value during the excitation frequency range of multiple solutions. As shown in Fig. 16, the region of multiple solutions will also change with the change of the nonlinear coefficient d . Our goal is to obtain the parameters of the nonlinear term in the system.

In the process of nonlinear system identification, it is necessary to first determine whether the response of the system is in the region of multiple solutions. Then a classifier is used to identify which solution branch is arrived under the influence of the current initial value. Based on the result, the corresponding identification model is

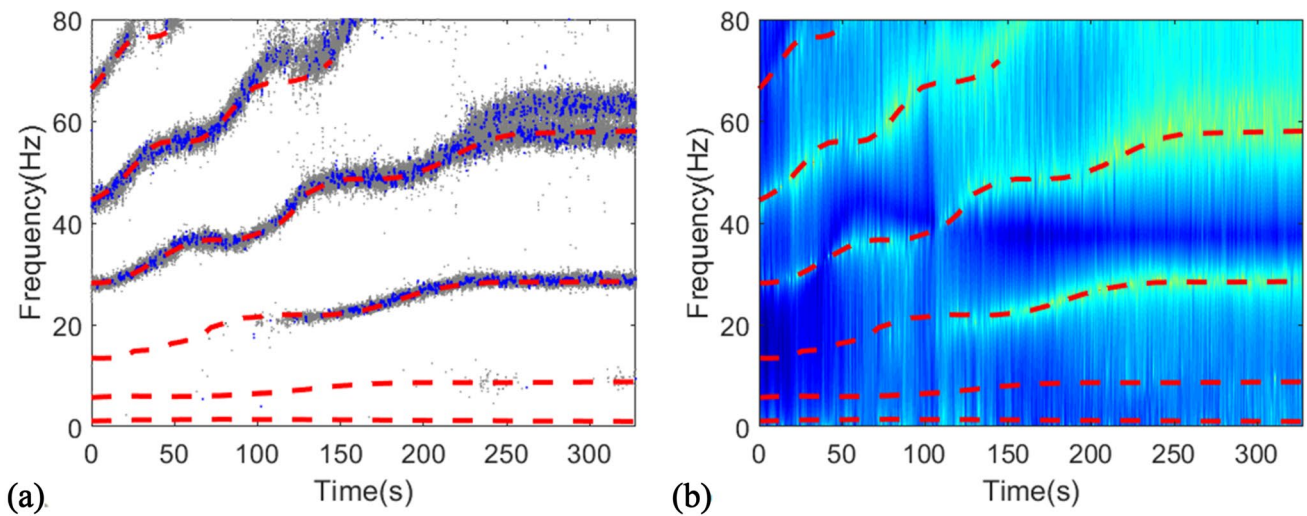


Fig. 14 Natural frequency and PSD estimates (gray dots: estimated natural frequencies from the 30 Monte Carlo tests; red dashed lines: baseline natural frequencies). (Color figure online)

Fig. 15 Framework of nonlinear system parameter identification

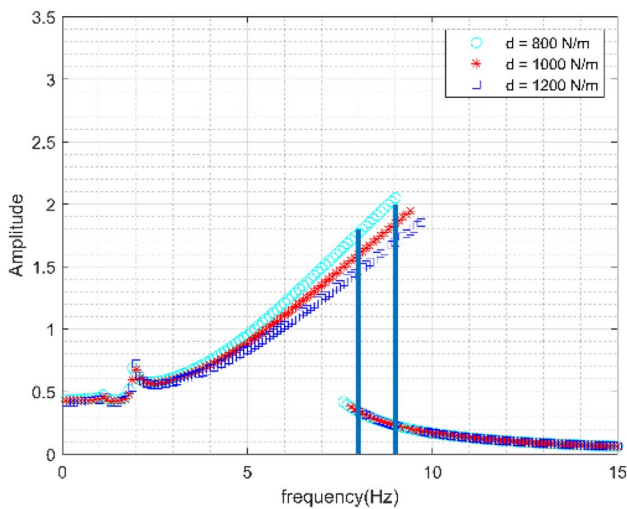
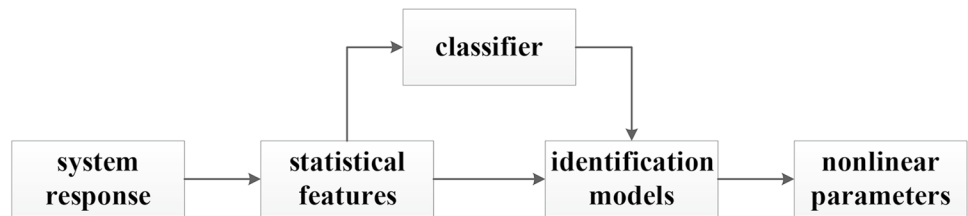


Fig. 16 The resonant amplitude-frequency curves change with the nonlinear parameter

adopted to determine the parameter values. As shown in Fig. 17, the nonlinear coefficient d can be correctly determined by using any branch responses in the region of multiple solutions.

4.3 Hybrid modeling of nonlinear jointed structures

Complex structural systems usually contain a number of connecting joints, in which nonlinear dynamic characteristics can be included, such as friction, freeplay, collision, etc. The mechanical parameters of the nonlinear joints are difficult to determine. Hence, a novel hybrid modeling method for nonlinear jointed structures based on finite element model reduction and DL techniques is proposed. As shown in Fig. 18, the main idea is summarized as follows: Firstly, finite element models of linear components are reduced to improve the computing efficiency through the free-interface mode synthesis method. Secondly, DNNs are used to equivalently represent the nonlinear joints which are difficult to describe by accurate and physically-motivated models. Nonlinear joints are finally replaced with their equivalent neural networks and connected with the substructure models of linear components through the compatibility of displacements and equilibrium of forces at the interfaces.

The problem of nonlinear system identification in its most general form is the construction of the mapping relations between the inputs and outputs of the system. In many circumstances accessing prior knowledge are very difficult, which makes the selection of an accurate and physically-motivated model for the system being identified virtually

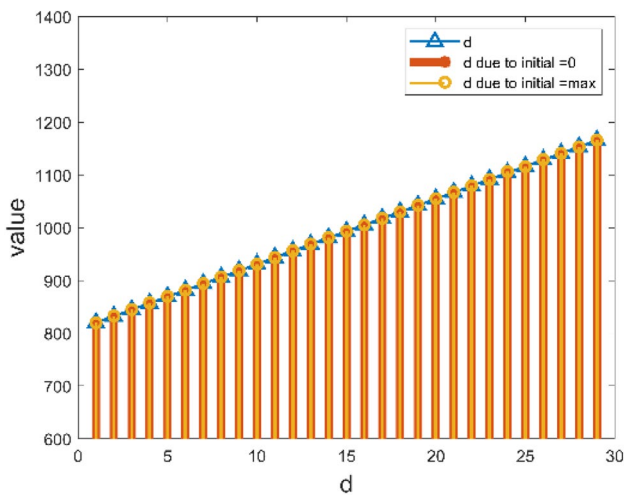


Fig. 17 The same system parameters corresponding to different responses

impossible. Neural network models are able to take advantage of a sufficiently rich and flexible mathematical structure to capture the underlying nonlinear physics in input–output data, without incorporating any prior knowledge [8]. Multiple experiments with different excitations are first conducted to acquire more accurate nonlinear characteristics of a joint. Force-state mapping of each experiment is established and a DNN is subsequently trained and validated based on a relatively larger number of mapping relations $(x, \dot{x}) \rightarrow f(x, \dot{x})$, with the states at both ends of the joints being the input and the restoring forces being the output of the network. Once the mapping relations $(x, \dot{x}) \rightarrow f(x, \dot{x})$ are captured and represented by the trained neural network, the hybrid model of a jointed structure can be built by replacing the nonlinear restoring force of each elastic connection with its equivalent neural network.

The performance of the hybrid modeling method is tested and assessed via a case study focused on a cantilever plate consisting of two linear components connected by three nonlinear joints, as shown in Fig. 19. Sinusoidal excitation in the translational direction of z-axis is applied to the nonlinear jointed structure with quadratic damping and freeplay stiffness in its hinge section, and responses of the hybrid

model are calculated by using numerical integration. Figure 20 shows the predicted displacement response and its fast Fourier transform (FFT) spectrum by the hybrid modeling method along with its true values. Evidently, both time and frequency characteristics of the displacement response predicted by the proposed hybrid modeling method are in good agreement with its true values, which demonstrates the response prediction capability of the proposed hybrid modeling method for nonlinear jointed structures. Besides, the appearance of harmonic components can be found in Fig. 20b due to the nonlinear joints, and the fact that only odd harmonics are present is a consequence of the restoring force function being odd.

5 Fault diagnosis and prognosis

In the past decades, the data-driven methods have been rapidly and successfully developed, which offer a promising tool for the industrial maintenance problems [51]. The data-driven methods generally aim to build relationships between the condition monitoring data and machinery health states, which are straight-forward and intuitive for the practical implementations. Basically, features can be firstly designed and extracted from the raw collected data, which are further used for evaluations of the machine health states afterwards.

As an emerging highly effective data-driven technology, DL has been successfully and popularly developed

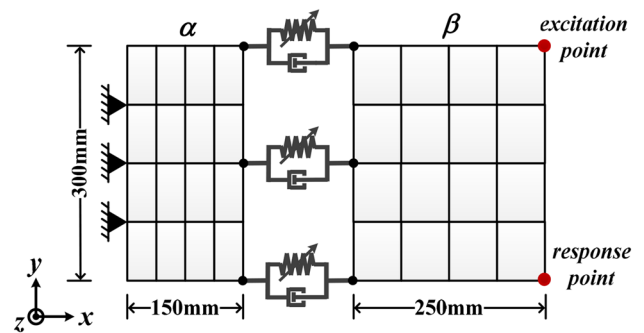


Fig. 19 Numerical model of a nonlinear cantilever plate

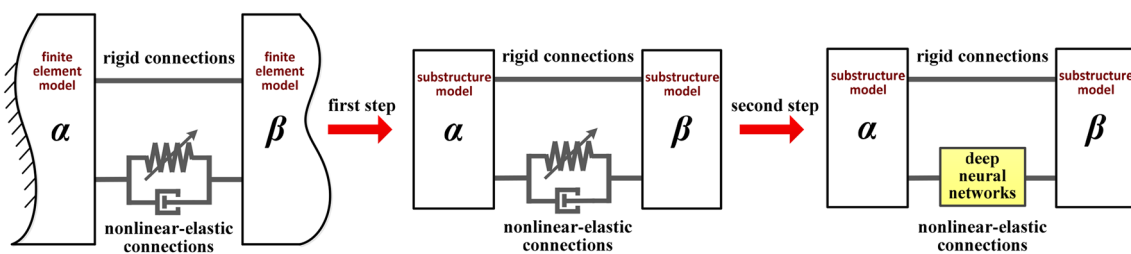


Fig. 18 Hybrid modeling of a jointed structure with rigid and nonlinear-elastic connections

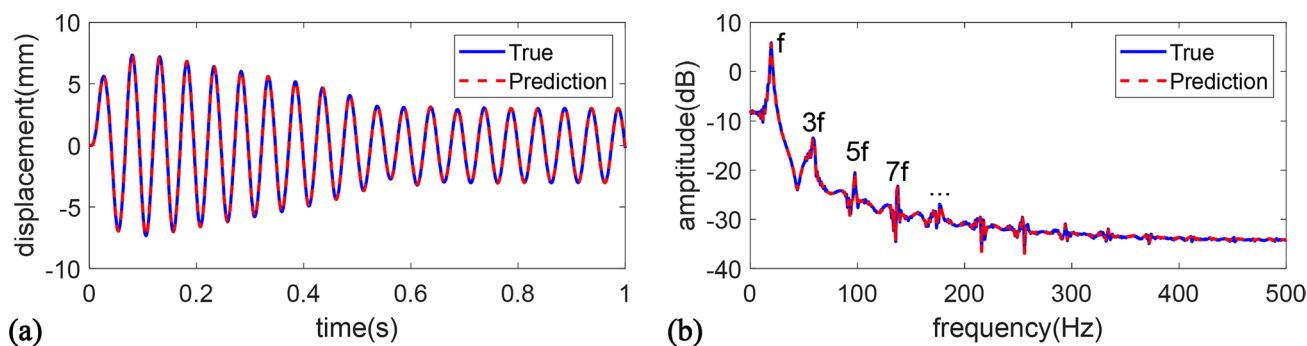


Fig. 20 Displacement response predicted by the hybrid modeling method and its true values under sinusoidal excitation: **a** displacement response and **b** its FFT spectrum

for fault diagnosis and prognosis problems in the literature. Through multiple linear and nonlinear transformations, different levels of data representations can be learned from the raw data, which can be further used in the down-stream tasks. As Fig. 21 shows, DL offers an intuitive end-to-end implementation scheme, which can directly process the raw measured data as model inputs and outputs the expected results such as machine health states. Less manually extracted features and little prior expertise on signal processing are basically required, which facilitates the applications in the real industries. More importantly, state-of-the-art fault diagnosis and prognosis results have been achieved by DL in the recent years, which show that DNN is well suited for the machinery maintenance problems.

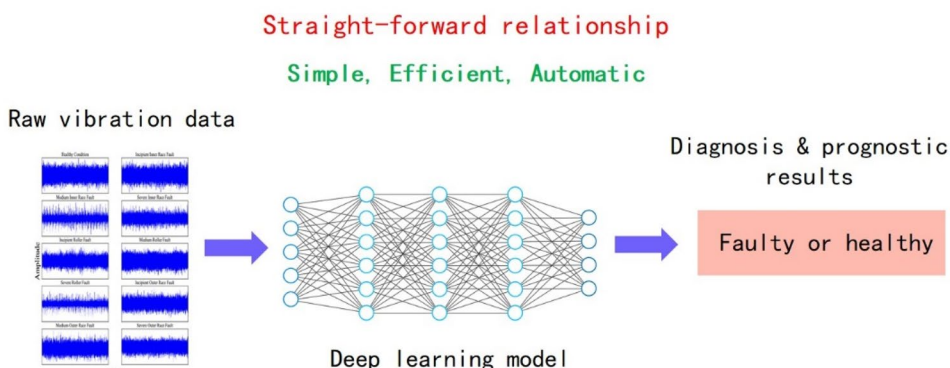
The following advances in data-driven fault diagnosis and prognosis methods cover a wide range of the specific problems and challenges in the industries. The concerned topics generally include the machinery health assessment, fault diagnosis, degradation prediction, remaining useful life estimation, transfer learning etc. In general, the latest DL techniques are focused on.

5.1 Fault diagnosis

The conventional fault diagnosis problem has been successfully addressed using the latest data-driven methods, where the training and testing data are from the same condition. A deep residual learning-based fault diagnosis method was proposed for rolling element bearings [52]. The deep CNN is used for automatic feature extraction, which is followed by the fully-connected layer for information aggregation. At the end of the network, soft-max function is adopted to interpret the obtained neuron values to predicted probabilities for different machine health conditions. To increase the model learning capability, the residual connections are proposed, which facilitate gradient propagation throughout the network, and reduces the risk of gradient vanishing. As a result, the learning performance is much improved with the deep residual structure.

Generally it requires large economic costs and much labor to collect sufficient accurately labeled training data with respect to different machine health conditions for the DL approaches. A data augmentation method was proposed aiming to enlarge the training dataset using limited data [53]. Multiple signal processing techniques are proposed to artificially create additional fake samples, including applying Gaussian noise, signal translation, amplitude shifts etc.

Fig. 21 End-to-end structure of DL-based fault diagnosis and prognosis framework



Besides the signal processing techniques for data augmentation, DL itself can also be developed to create additional samples for better training. A deep generative adversarial network-based method was proposed for the data unbalance problem [54]. Comparing to the data in healthy state, the data in faulty state of machinery are difficult to collect. One of the most straight-forward solutions lies in balancing the unbalanced dataset. In that study, a generative model is utilized to learn the mapping which projects noise to the real data. In this way, additional fake samples can be generated, and the training dataset can be sufficiently enlarged.

With respect to the limited labeled training data problem, while creating additional fake data can effectively enlarge the valid dataset, it is still difficult to guarantee the diversity of the generated samples. In many cases, noticeable high level of similarity is observed from the fake data, which bring negative influence for the model training performance. Instead of creating fake samples, we proposed to utilize the real-world unlabeled data to assist model training with limited labeled data [55]. A deep representation clustering algorithm is adopted, where a pre-training stage is implemented to firstly learn the high-level features from raw data. Next, clustering of the unlabeled data is carried out in the new sub-space, and pseudo labels are attached. Therefore, the unlabeled data can be further used for supervised training to improve model performance.

Despite the promising results achieved by DL in fault diagnosis, the neural network model generally remains a “black-box” to users, and the internal mechanism of DL has been less understood. The interpretability of DL is supposed to be enhanced, and the study in [56] offers one of the first attempts in discovering the mechanism of neural network in fault diagnosis. An attention module is proposed to obtain the weights attached by the model to different data segments. The importance of different input segments on the diagnosis results can be thus achieved. It is observed in the experiments that DL also automatically learns the conventional features for diagnosis, such as the fault characteristic frequencies etc., rather than overfitting the data.

5.2 Prognosis

Machinery prognostic tasks are generally more challenging than the fault diagnosis problems, since it is of high uncertainty to predict the future health conditions during operation. In [57], a deep convolutional neural network model was introduced to the remaining useful life (RUL) prediction problem. The time series condition monitoring data are segmented into multiple samples as model inputs. Through multiple convolutional and pooling operations, useful prognostic features can be automatically learned, which can be further used for RUL estimation.

In the current literature, the conventional deep convolutional neural network generally adopts the sequential data processing scheme [58]. The higher-level features are learned through multiple layers, which are used for the down-stream tasks, while the knowledge learned from the low-level features are discarded. That results in information losses and has higher risk of overfitting. A multi-scale deep convolutional neural network was proposed for RUL prediction, where different levels of features in multiple layers are concatenated [59]. In this way, both the low- and high-level features are utilized for prognosis, which enhances the learning capacity of the network. Experiments on the popular accelerated aging platform PRONOSTIA for rolling element bearings are carried out for validations.

Generally, machine operates in healthy state in the early period. At a certain time step, an initial fault may occur such as a small crack, and then the machine starts to degrade afterwards. Therefore, it is important to locate the degradation starting point in data-driven prognostic studies, which is also popularly known as the first predicting time (FPT). However, the incipient fault is difficult to identify in condition monitoring, and the proper determination of FPT is thus quite challenging. To address this issue, a DL-based prognostic method is proposed in [60], where the GAN is used to learn the distributions of data in machine healthy states. A metric in the GAN framework can readily and effectively serve as the health indicator for degradation, which thus well indicates the FPT.

At present, the condition monitoring data such as vibration acceleration, current signals etc. have been popularly used for prognostics. It is noted that in the intelligent maintenance problems, the image-based machine vision techniques are also promising for identifying degradation [61, 62]. The work in [63] contributes efforts on proposing a new prognostic framework based on the component images collected in the real production lines. Specifically, the DL scheme is utilized taking the images as the inputs, which directly outputs the RUL of the cutting wheels. As a special focus in that study, partial observation of the component is considered, due to the practical limitations in the mechanical structure. Correspondingly, a supervised attention mechanism is proposed in the DL framework to automatically focus on the informative data for prognostics and ignore the non-discriminative data. The proposed method well addressed the challenging RUL prediction problem with image data incompleteness and disturbances.

5.3 Transfer learning across operating conditions

For DL-based data-driven tasks, one of the main assumptions lies in that the training and testing data are supposed to be from the same distribution, i.e., the data should be collected from the same machine under identical operating

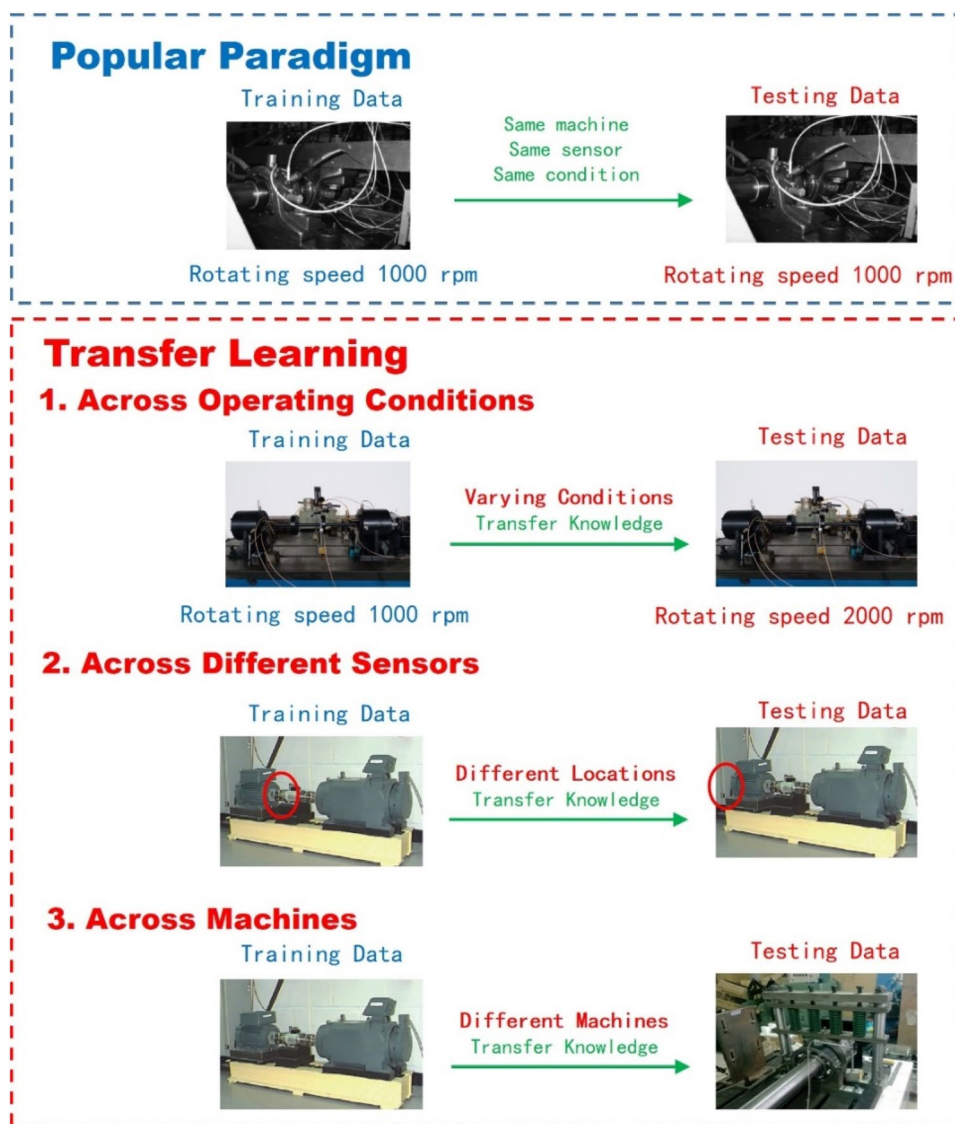
conditions. In the real industries, different types of machines are usually used in practice, and different machines are also under different operating regimes. That is the training and testing data are mostly from different distributions [64]. That poses great obstacles in applying the learned fault diagnosis knowledge from the training data, denoted as source domain, on the testing data, denoted as target domain. In order to bridge the domain gap, transfer learning techniques have been developed [65]. Especially, domain adaptation approaches have been popularly developed in transfer learning for fault diagnosis [66, 67]. Figure 22 shows different types of transfer learning in machine fault diagnosis problems.

A DL-based domain adaptation method was proposed in [68] for fault diagnosis, where different operating conditions are considered for the training and testing data. In order to minimize the gaps between the source and target domain

samples, their distances in the high-level representation space are measured and optimized using the maximum mean discrepancy (MMD) metric. The MMDs in multiple layers including the convolutional and fully-connected layers are minimized for the model to extract the domain-invariant features. In this way, generalized fault diagnosis knowledge can be learned, which facilitates knowledge transfer across different operating conditions. An adversarial multi-classifier optimization method for cross-domain fault diagnosis was proposed to achieve class-level domain adaptation effects [69].

Different levels of environmental noise may also exist in the training and testing data. That could lead to ineffectiveness of the data-driven model in practical implementations. A distance metric learning approach is proved to be able to further enhance the generalization ability of the fault diagnosis model [70], in which the intra-class distances are

Fig. 22 Different transfer learning schemes in machine fault diagnosis problems



minimized with respect to different health conditions and the inter-class distances are enlarged to maximize the distances between different states. The learned features in the high-level layers can be more generalized and insensitive to the noise disturbances.

One of the main drawbacks in the existing domain adaptation methods lies in the assumption of availability of testing data during model training. That facilitates explicit knowledge transfer for data across domains. However, the testing data cannot be obtained in advance in most real cases. In [71], a domain augmentation technique is first applied to enlarge the available training dataset, in order to simulate different operating conditions. Adversarial training is then implemented on the augmented domains to extract generalized features. The learned features thus can be also applied on new domains where data are not available in training. Distance metric learning is further integrated to increase the model robustness.

With respect to the practical data unavailability problem in the transfer learning tasks, a generative neural network model was proposed in [72], aiming to learn a mapping function, which projects the healthy state data to the faulty state data in the high-level layers. Therefore, in testing scenarios, fake faulty state data can be artificially created using the healthy state data, and the conventional domain adaptation approaches can be thus readily used for cross-domain fault diagnosis. In addition, the label space of the target domain is generally contained in that of the source domain in most cases, which brings great challenge in domain adaptation [73]. A partial domain adaptation method was proposed in [74], where the unlabeled target-domain training data do not cover the full machinery health condition label space. Conditional data alignment between source and target domains is proposed to achieve class-level adaptation, and unsupervised prediction consistency scheme is introduced [75]. Taking advantages of multiple classification modules, promising partial transfer learning performance has been achieved.

5.4 Transfer learning across sensors and machines

Most existing studies assume that the data are collected from the same sensors during training and testing. However, sensor malfunction could possibly occur. The data-driven model will inevitably lose effectiveness if the testing data cannot be obtained from the same sensors from which the training data are measured. Compared with the gap between operating conditions, the data from different sensors are more heterogeneous. That makes the cross-sensor transfer learning problem very challenging. In [76], the feature extraction module is used to obtain the high-level representations of different sensor data, and the domain discriminator is adopted to distinguish the source of the data. Through adversarial training between the feature extractor and domain discriminator,

generalized features can be learned across sensors. Furthermore, parallel data can be easily collected from different sensors, and the alignments of them in the hidden layers help to improve the transfer performance.

For the cases that the training and testing data are collected from different machines, installation condition, sampling frequency etc., it is quite difficult to transfer the fault diagnosis knowledge across different machines. A deep domain adaptation method was proposed for cross-machine fault diagnosis problems [77]. Auto-encoder architecture is firstly used to project data from different machines into the same data sub-space, which facilitates further knowledge transfer. After pre-training, the MMD metric is adopted, the minimization of which draws the data from different machines into the same regions.

Focusing on the cross-machine transfer learning problems, a more generalized model was built in [78], where the data from multiple machines are explored, and machine-invariant features are extracted for diagnosing the testing machine with limited training data. Specifically, multiple source machines are exploited, which share similar but different types. An adversarial learning scheme is proposed. Besides the feature extractor and domain discriminator, multiple classifiers are used corresponding with multiple sources. Through adversarial training with a wide range of data, machine-invariant features can be learned and benefit the model establishment of the target machine that has limited training data.

6 Heart rate variability analysis

Heart rate variability (HRV) analysis has been widely used to evaluate autonomic nervous system activity [79], so as to carry out objective evaluations of noxious stimulation. However, the significance and meaning of many different measures of HRV are more complex than generally appreciated and there is a potential for incorrect conclusions and for excessive or unfounded extrapolations [80, 81]. There is evidence that physiological signals under healthy conditions may have a fractal temporal structure [82]. Entropy-based measures, such as the typical approximate entropy and sample entropy, have been widely used in HRV analysis [83].

ML is used to predict the clinical pain by using multimodal neuroimaging and autonomic metrics, the BIS during target-controlled infusion of propofol and remifentanyl, and hypotension based on high-fidelity arterial pressure waveform analysis [84, 85]. In HRV signal analysis, Ong et al. [86] found that ML scores are more accurate than the modified early warning score in predicting cardiac arrest within 72 h. The long-term ambulatory index may have applications in the study and management of other mental illnesses, and it is useful for other

clinical disciplines where cardiovascular disease and stress are significant factors. Chiew et al. [87] found that a ML model incorporating HRV analysis can be used to improve prediction of 30-day in-hospital mortality among suspected sepsis patients in the emergency department compared to traditional risk stratification tools. These methods may assist clinicians in the faster diagnosis of the patients and to provide timely treatment.

6.1 Data collection and data preparation

Sixty patients have undergone oral and maxillofacial surgery under the general anesthesia. Patients' ECG signals are continuously recorded by BMD101 (NeuroSky Inc.) at a sampling rate of 512 Hz and stored in the computer during peri-operation. Signal processing is carried out to remove noise and ectopic rhythm, identify the R wave of each cardiac cycle, and form the RR interval time series to be analyzed. Three kinds of signals are taken from the complete RR interval sequence: preoperative (T0), intubation (T1) and intraoperative (T2).

The data consists of a set of RR interval signals and is divided into three different classes: the consciousness, the general anesthesia and the tracheal intubation stimulation. The data of selected cases were randomly allocated to three sets of data: training, validation, and testing sets. Data are randomly split into training and test data at a 70:15:15 ratio, with the class balance in training, validation and test sets maintained to reflect the class balance in the entire data set. Both the training and validation sets are used for modeling. The test set is the remaining cases that is not used for modeling but is used to test the performance of the final model.

6.2 Data processing

The empirical mode decomposition method is developed based on a hypothesis that any signal consists of various simple intrinsic mode functions (IMF). The procedure includes confirming all the local extrema as the upper envelope, calculating the means of the upper envelope and the lower envelope as the trend curve, and computing the difference from the original signal and the trend curve as the detrending curve. One example for RR interval signal decomposed into a trend curve and a detrending curve is shown in Fig. 23.

There are two steps in classification of patients in consciousness or under the general anesthesia by using DNNs. Step 1, the sorted database is processed to classify the consciousness and the general anesthesia. Step 2, the data under the general anesthesia is processed to classify whether the tracheal intubation stimulation occurs under the general anesthesia. The DNN model consists of the input layer, the LSTM layer, the fully connected layer and classification output layer. During training, the weights of the networks are calculated with ADAM optimizer. The optimal model without overfitting is selected by using training and validation sets. The final model is applied to the test set for testing the performance of the model.

6.3 Case studies

During the study period, the signals are from 60 patients and a total of 182 signals are selected after screening. Signals are randomly divided into training, validation and test data sets with ten times repetition. As shown in Fig. 24, one finds that the accuracy of classification results of the proposed classification strategy is 96.55%, 96.55%, 93.10%,

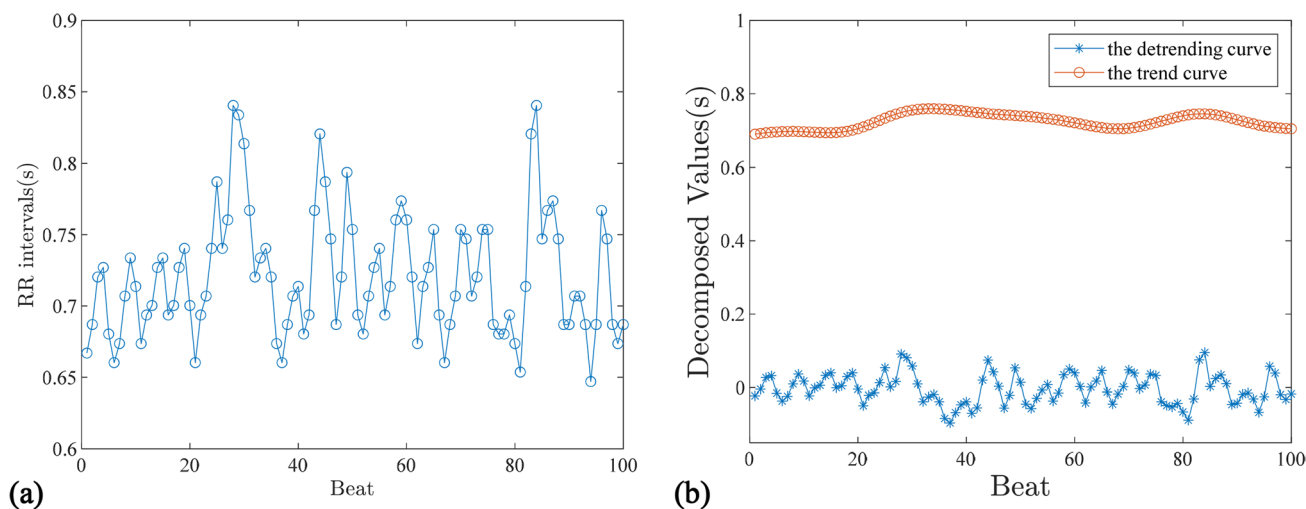


Fig. 23 a RR interval signal and b a trend curve and a detrending curve

93.10%, 96.55%, 86.21%, 96.55%, 93.10%, 89.66% and 93.10%, while the accuracy of classification results that use directly the RR interval signals as input is 44.83%, 51.72%, 51.72%, 51.72%, 51.72%, 58.62%, 55.17%, 48.28%, 58.62% and 37.93%. This means that the feature extraction in the data processing is valuable and the features extracted are consistent with the physiological significance. The average classification accuracy based on these results with ten times repetition is shown in Fig. 25. The correct total numbers of test data sets of the direct classification are 148/290. The numbers of the consciousness, the general anesthesia and the tracheal intubation stimulation are 62/120, 60/100 and 26/70. The correct total numbers of test data sets of the proposed classification strategy are 271/290. The numbers of

the consciousness, the general anesthesia and the tracheal intubation stimulation are 117/120, 91/100 and 63/70. The values of the average accuracy of the direct classification are 51.03%, 51.67%, 60.00% and 37.14%. The values of the average accuracy of the proposed classification strategy are 93.45%, 97.50%, 91.00% and 90.00%. Evidently, the average accuracy is significantly improved by using the proposed classification strategy.

7 Summary and future directions

In this paper, we have presented an overview of our group’s recent research work based on data-driven methods in the field of dynamics and control, including structural optimization, active vibration control, system identification, fault diagnosis and prognosis, and state identification of HRV signal. Some potential future research directions and trends are discussed next.

Structural optimization many state-of-the-art data-driven methods have been recently proposed to improve the efficiency of structural optimization, such as multi-fidelity modeling, deep reinforcement learning and data-driven evolutionary optimization. There are still some challenges in this filed, such as lack of data and surrogate model reliability. Therefore, future work includes the development of surrogate models combined with physical mechanisms and the application of transfer learning and ensemble learning in the optimization process.

Active vibration control the data-driven ADP method is promising to design active vibration controllers, which directly use on-line input and output data without need of explicit information of the mathematical model. For such data-driven control method, how to give the initial stable control law and

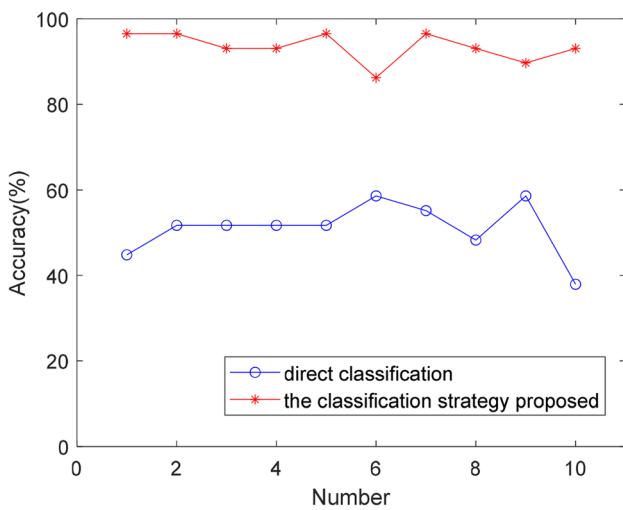


Fig. 24 The accuracy of the classification strategy proposed and direct classification

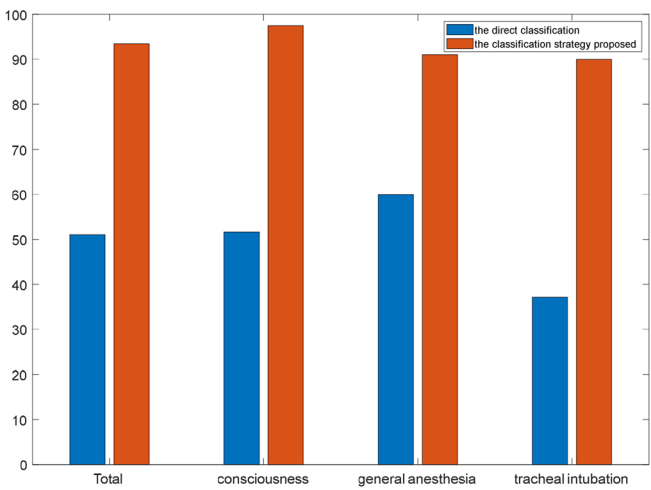
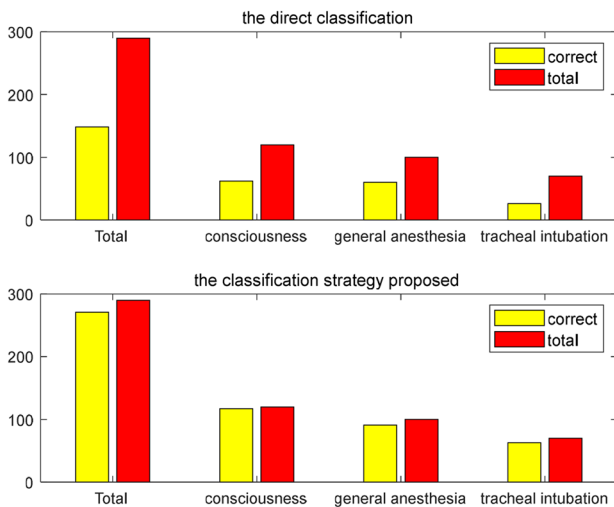


Fig. 25 The average classification accuracy

handle the collected data with sensor noise to ensure the convergence remains to be further investigated.

System identification it is still of great significance to develop enhanced system identification methods based on the recent theories of ML and DL, such as estimators with stronger anti-noise ability for time-varying/nonlinear systems, parameter identification of high-dimensional nonlinear systems, DL-assisted data preprocessing and model structure selection. Hybrid modeling based on the combination of the physically-motivated model and the data-driven model is worth more investigation to solve the dynamic problems of large-scale complex structures. Besides, there also exists an urgent need to investigate the changes of system parameters and monitor the health of the system through key parameter identification.

Fault diagnosis and prognosis despite the promising prospects, some challenges of DL in fault diagnosis and prognosis still exist. *Interpretability* the internal mechanism of data processing in DNNs has not been clear to researchers and practitioners. Further investigation of the interpretability can increase the applicability of DL. *Data availability* model establishment generally requires a large amount of labeled training data. But, collecting high-quality and accurately labeled data is always difficult and expensive in the real industries. Further researches may focus on the explorations of limited labeled training data for DL. *Data privacy* collaboration of multiple users for collecting sufficient data for training a global maintenance model is promising in practice. Considering potential conflict of interests, the data privacy-preserving strategies is very important.

Heart rate variability analysis the features of RR interval signals can be extracted to successfully evaluate the activity of the autonomic nervous system by HRV analysis. It should be noted that the diagnoses and monitoring in clinical medicine rely heavily on various biological signals from the human body. The DL-based application in the medical field is very promising. Quantitative evaluation of the noxious stimulation's intensity is more valuable in depth of anesthesia monitoring, which is worth much more investigation.

Acknowledgements This work was supported by the National Natural Science Foundation of China (Grant Nos. 11972245, 11802201) and the Natural Science Foundation of Tianjin City (Grant No. 19JCZDJC38800).

Compliance with ethical standards

Conflict of interest The authors declared that they have no conflict of interest to this work.

References

- Zhang S, Zhang S, Wang B et al (2020) Machine learning and deep learning algorithms for bearing fault diagnostics—a comprehensive review. *IEEE Access* 8:29857–29881
- Alom MZ, Taha T, Yakopcic C et al (2018) The history began from AlexNet: a comprehensive survey on deep learning approaches. *arXiv preprint arXiv:1803.01164*
- NG A (2018) CS229 course notes: deep learning. Stanford University
- Prabhat (2017) A look at deep learning for science—scientific use cases show promise, but challenges remain for complex data analytics. O'Reilly Media
- Lagaris IE, Likas A, Fotiadis DI (1998) Artificial neural networks for solving ordinary and partial differential equations. *IEEE Trans Neural Netw* 9:987–1000
- Chen RTQ, Rubanova Y, Bettencourt J et al (2019) Neural ordinary differential equations. *arXiv:1806.07366*
- Wei S, Jin X, Li H (2019) General solutions for nonlinear differential equations: a rule-based self-learning approach using deep reinforcement learning. *Comput Mech* 64:1361–1374
- Noël JP, Kerschen G (2017) Nonlinear system identification in structural dynamics: 10 more years of progress. *Mech Syst Signal Process* 83:2–35
- Zhao R, Yan R, Chen Z et al (2019) Deep learning and its applications to machine health monitoring. *Mech Syst Signal Process* 115:213–237
- Naranjani Y, Hernández C, Xiong F-R et al (2017) A hybrid method of evolutionary algorithm and simple cell mapping for multi-objective optimization problems. *Int J Dyn Control* 5:570–582
- Zitzler E, Thiele L (1998) Multiobjective optimization using evolutionary algorithms—a comparative case study. In: *Parallel problem solving from nature—PPSN V*. Springer, Berlin Heidelberg, pp 292–301
- Deb K, Pratap A, Agarwal S et al (2002) A fast and elitist multiobjective genetic algorithm: {NSGA-II}. *IEEE Trans Evol Comput* 6:182–197
- Coello CAC, Pulido GT, Lechuga MS (2004) Handling multiple objectives with particle swarm optimization. *IEEE Trans Evol Comput* 8:256–279
- Zhang W, Han B, Li X et al (2020) Multi-objective system optimization method and experimental validation of a centralized squeeze film damper using a cell mapping method considering dynamic constraints. *Eng Optim*. <https://doi.org/10.1080/0305215X.2020.1759577>
- Zhang W, Han B, Li X et al (2019) Multiple-objective design optimization of squirrel cage for squeeze film damper by using cell mapping method and experimental validation. *Mech Mach Theory* 132:66–79
- Li X, Sun J-Q (2019) Multi-objective optimal predictive control of signals in urban traffic network. *J Intell Transp Syst* 23:370–388
- Li X, Sun J (2018) Signal multiobjective optimization for urban traffic network. *IEEE Trans Intell Transp Syst* 19:3529–3537
- Li X, Sun J-Q (2019) Turning-lane and signal optimization at intersections with multiple objectives. *Eng Optim* 51:484–502
- Li X, Sun J-Q (2019) Intersection multi-objective optimization on signal setting and lane assignment. *Phys A* 525:1233–1246
- Jin Y, Wang H, Chugh T et al (2018) Data-driven evolutionary optimization: an overview and case studies. *IEEE Trans Evol Comput* 23:442–458
- Cortes C, Vapnik V (1995) Support-vector networks. *Mach Learn* 20:273–297
- Elmeligy A, Mehrani P, Thibault J (2018) Artificial neural networks as metamodels for the multiobjective optimization of biobutanol production. *Appl Sci* 8:961

23. Martinez SZ, Coello CAC (2013) MOEA/D assisted by rbf networks for expensive multi-objective optimization problems. Paper presented at the Proceedings of the 15th annual conference on Genetic and evolutionary computation, Amsterdam, The Netherlands
24. Zhang Q, Liu W, Tsang E et al (2010) Expensive multiobjective optimization by MOEA/D with Gaussian process model. *IEEE Trans Evol Comput* 14:456–474
25. He M-X, Xiong F-R, Sun J-Q (2017) Multi-objective optimization of elastic beams for noise reduction. *J Vib Acoust* 139:051010–051014
26. He M-X, Sun J-Q (2018) Multi-objective structural-acoustic optimization of beams made of functionally graded materials. *Compos Struct* 185:221–228
27. Zhao C, Prasad MG (2019) Acoustic black holes in structural design for vibration and noise control. *Acoustics* 1:220–251
28. Lyu X, Ding Q, Yang T (2020) Merging phononic crystals and acoustic black holes. *Appl Math Mech* 41:279–288
29. Li X, Ding Q (2018) Analysis on vibration energy concentration of the one-dimensional wedge-shaped acoustic black hole structure. *J Intell Mater Syst Struct* 29:2137–2148
30. Li X, Ding Q (2019) Sound radiation of a beam with a wedge-shaped edge embedding acoustic black hole feature. *J Sound Vib* 439:287–299
31. Tang L, Cheng L (2017) Broadband locally resonant band gaps in periodic beam structures with embedded acoustic black holes. *J Appl Phys* 121:194901
32. Lewis FL, Vrabie D, Syrmos VL (2012) *Optimal control*. Wiley, Hoboken
33. Werbos P (1992) Approximate dynamic programming for realtime control and neural modelling. In: *Handbook of intelligent control: neural, fuzzy and adaptive approaches*, pp 493–525
34. Vrabie D, Lewis F (2009) Neural network approach to continuous-time direct adaptive optimal control for partially unknown nonlinear systems. *Neural Netw* 22:237–246
35. Yang X, He H, Zhong X (2017) Adaptive dynamic programming for robust regulation and its application to power systems. *IEEE Trans Ind Electron* 65:5722–5732
36. Liu D, Javaherian H, Kovalenko O et al (2008) Adaptive critic learning techniques for engine torque and air–fuel ratio control. *IEEE Trans Syst Man Cybern Part B (Cybern)* 38:988–993
37. Werbos PJ (1992) Approximate dynamic programming for real-time control and neural modelling. *Handbook of intelligent control: neural, fuzzy and adaptive approaches*, pp 493–525
38. Xue W-T, Li J-Z (2016) An adaptive backstepping sliding mode control method for aeroelastic system. In: *IECON 2016-42nd annual conference of the IEEE Industrial Electronics Society*. IEEE, pp 6764–6769
39. Kerschen G, Worden K, Vakakis AF et al (2006) Past, present and future of nonlinear system identification in structural dynamics. *Mech Syst Signal Process* 20:505–592
40. Poulimenos AG, Fassois SD (2006) Parametric time-domain methods for non-stationary random vibration modelling and analysis—a critical survey and comparison. *Mech Syst Signal Process* 20:763–816
41. Spiridonakos MD, Fassois SD (2014) Non-stationary random vibration modelling and analysis via functional series time-dependent ARMA (FS-TARMA) models—a critical survey. *Mech Syst Signal Process* 47:175–224
42. Ma Z, Ding Q, Liu L et al (2018) Research progress on time-domain modal parameter estimation methods for linear time-varying structures. *J Mech Eng* 54:137–159
43. Ma Z-S, Liu L, Zhou S-D et al (2018) Parametric output-only identification of time-varying structures using a kernel recursive extended least squares TARMA approach. *Mech Syst Signal Process* 98:684–701
44. Ma Z-S, Ding Q, Zhou S-D (2020) Novel adaptive methods for output-only recursive identification of time-varying systems subject to gross errors. *J Vib Control* 26:306–317
45. Ma Z-S, Ding Q, Tang Y (2020) Operational modal analysis of a liquid-filled cylindrical structure with decreasing filling mass by multivariate stochastic parameter evolution methods. *Int J Mech Sci* 172:105420
46. Ma Z-S, Liu L, Zhou S-D et al (2016) Output-only recursive identification of time-varying structures using a Gaussian process regression TARMA approach. In: *ISMA2016 international conference on noise and vibration engineering*, Leuven, 19–21 September 2016, pp 2859–2872
47. Ma Z-S, Li L, Ding Q (2020) Multivariate recursive Bayesian linear regression and its applications to output-only identification of time-varying. *J Vib Control*. <https://doi.org/10.1177/1077546320941703>
48. Ma Z-S, Liu L, Zhou S-D et al (2017) Output-only modal parameter regression estimation of time-varying structures via a kernel ridge regression FS-TARMA approach. *Shock Vib* 2017:8176593
49. Zhou S-D, Ma Y-C, Liu L et al (2018) Output-only modal parameter estimator of linear time-varying structural systems based on vector TAR model and least squares support vector machine. *Mech Syst Signal Process* 98:722–755
50. Ma C, Li W, Cao J et al (2020) Adaptive sliding window based activity recognition for assisted livings. *Inf Fusion* 53:55–65
51. Yang Q, Li X, Wang Y et al (2020) Fault diagnosis of ball screw in industrial robots using non-stationary motor current signals. *Procedia Manuf* 48:1102–1108
52. Zhang W, Li X, Ding Q (2019) Deep residual learning-based fault diagnosis method for rotating machinery. *ISA Trans* 95:295–305
53. Li X, Zhang W, Ding Q et al (2020) Intelligent rotating machinery fault diagnosis based on deep learning using data augmentation. *J Intell Manuf* 31:433–452
54. Zhang W, Li X, Jia X-D et al (2020) Machinery fault diagnosis with imbalanced data using deep generative adversarial networks. *Measurement* 152:107377
55. Li X, Li X, Ma H (2020) Deep representation clustering-based fault diagnosis method with unsupervised data applied to rotating machinery. *Mech Syst Signal Process* 143:106825
56. Li X, Zhang W, Ding Q (2019) Understanding and improving deep learning-based rolling bearing fault diagnosis with attention mechanism. *Signal Process* 161:136–154
57. Li X, Ding Q, Sun J-Q (2018) Remaining useful life estimation in prognostics using deep convolution neural networks. *Reliab Eng Syst Saf* 172:1–11
58. Zhang W, Li X, Li X (2020) Deep learning-based prognostic approach for lithium-ion batteries with adaptive time-series prediction and on-line validation. *Measurement* 164:108052
59. Li X, Zhang W, Ding Q (2019) Deep learning-based remaining useful life estimation of bearings using multi-scale feature extraction. *Reliab Eng Syst Saf* 182:208–218
60. Li X, Zhang W, Ma H et al (2020) Data alignments in machinery remaining useful life prediction using deep adversarial neural networks. *Knowl-Based Syst* 197:105843
61. Yang S, Li X, Jia X et al (2020) Deep learning-based intelligent defect detection of cutting wheels with industrial images in manufacturing. *Procedia Manuf* 48:902–907
62. Li X, Siahpour S, Lee J et al (2020) Deep learning-based intelligent process monitoring of directed energy deposition in additive manufacturing with thermal images. *Procedia Manuf* 48:643–649
63. Li X, Jia X, Wang Y et al (2020) Industrial remaining useful life prediction by partial observation using deep learning with supervised attention. *IEEE/ASME Trans Mechatron*. <https://doi.org/10.1109/TMECH.2020.2992331>

64. Ainapure A, Li X, Singh J et al (2020) Enhancing intelligent cross-domain fault diagnosis performance on rotating machines with noisy health labels. *Procedia Manuf* 48:940–946
65. Xu N-X, Li X (2019) Cross-domain machinery fault diagnosis using adversarial network with conditional alignments. In: *Prognostics and System Health Management conference (PHM-Qingdao)*, pp 1–5
66. Ainapure A, Li X, Singh J et al (2020) Deep learning-based cross-machine health identification method for vacuum pumps with domain adaptation. *Procedia Manuf* 48:1088–1093
67. Azamfar M, Li X, Lee J (2020) Intelligent ball screw fault diagnosis using a deep domain adaptation methodology. *Mech Mach Theory* 151:103932
68. Li X, Zhang W, Ding Q et al (2019) Multi-Layer domain adaptation method for rolling bearing fault diagnosis. *Sig Process* 157:180–197
69. Li X, Zhang W, Ma H et al (2020) Deep learning-based adversarial multi-classifier optimization for cross-domain machinery fault diagnostics. *J Manuf Syst* 55:334–347
70. Li X, Zhang W, Ding Q (2018) A robust intelligent fault diagnosis method for rolling element bearings based on deep distance metric learning. *Neurocomputing* 310:77–95
71. Li X, Zhang W, Ma H et al (2020) Domain generalization in rotating machinery fault diagnostics using deep neural networks. *Neurocomputing* 403:409–420
72. Li X, Zhang W, Ding Q (2019) Cross-domain fault diagnosis of rolling element bearings using deep generative neural networks. *IEEE Trans Ind Electron* 66:5525–5534
73. Li X, Zhang W, Ma H et al (2020) Partial transfer learning in machinery cross-domain fault diagnostics using class-weighted adversarial networks. *Neural Netw* 129:313–322
74. Li X, Zhang W (2020) Deep learning-based partial domain adaptation method on intelligent machinery fault diagnostics. *IEEE Trans Ind Electron*. <https://doi.org/10.1109/TIE.2020.2984968:1-1>
75. Li X, Jia X, Yang Q et al (2020) Quality analysis in metal additive manufacturing with deep learning. *J Intell Manuf*. <https://doi.org/10.1007/s10845-020-01549-2>
76. Li X, Zhang W, Xu N-X et al (2019) Deep learning-based machinery fault diagnostics with domain adaptation across sensors at different places. *IEEE Trans Ind Electron* 67:6785–6794
77. Li X, Jia X-D, Zhang W et al (2020) Intelligent cross-machine fault diagnosis approach with deep auto-encoder and domain adaptation. *Neurocomputing* 383:235–247
78. Li X, Zhang W, Ding Q et al (2019) Diagnosing rotating machines with weakly supervised data using deep transfer learning. *IEEE Trans Ind Inf* 16:1688–1697
79. Kleiger RE, Stein PK, Bigger JT (2005) Heart rate variability: measurement and clinical utility. *Ann Noninvasive Electrocardiol* 10:88–101
80. Yang ACC, Hseu SS, Yien HW et al (2003) Linguistic analysis of the human heartbeat using frequency and rank order statistics. *Phys Rev Lett* 90:108103
81. Kraskov A, Nadler W, Stogbauer H et al (2004) Comment on “Linguistic analysis of the human heartbeat using frequency and rank order statistics”. *Phys Rev Lett* 92:109801
82. Ivanov PC, Amaral LA, Goldberger AL et al (1999) Multifractality in human heartbeat dynamics. *Nature* 399:461–465
83. Richman JS, Moorman JR (2000) Physiological time-series analysis using approximate entropy and sample entropy. *Am J Physiol Heart Circ Physiol* 278:2039–2049
84. Lee HC, Ryu HG, Chung EJ et al (2018) Prediction of bispectral index during target-controlled infusion of propofol and remifentanyl a deep learning approach. *Anesthesiology* 128:492–501
85. Hatib F, Jian ZP, Buddi S et al (2018) Machine-learning algorithm to predict hypotension based on high-fidelity arterial pressure waveform analysis. *Anesthesiology* 129:663–674
86. Ong MEH, Ng CHL, Goh K et al (2012) Prediction of cardiac arrest in critically ill patients presenting to the emergency department using a machine learning score incorporating heart rate variability compared with the modified early warning score. *Crit Care* 16:1–12
87. Chiew CJ, Liu N, Tagami T et al (2019) Heart rate variability based machine learning models for risk prediction of suspected sepsis patients in the emergency department. *Medicine* 98:e14197