Original Article

Anomaly detection based on multivariate data for the aircraft hydraulic system

Hongsheng Yan(iD), **Jianzhong Sun**(iD) **and Hongfu Zuo**

Abstract

It is almost impossible to detect the health status of the aircraft hydraulic system via a single variable, because of the complexity and the coupling relationship between components of the system. To serve the purpose, a novel anomaly detection method considering multivariate monitoring data is proposed in this article. The unsupervised auto-encoder model with the long short-term memory layers is used to reconstruct multivariate time series data, and a new comprehensive decision-making index based on two conventional ones is proposed to measure the difference between the observation and the reconstruction. Then, the health threshold of the decision-making index can be calculated by the kernel density estimation. The flight data are divided into several samples, and the anomaly detection of each sample is determined by the specific rule. The healthy status of each flight is determined by voting based on the detection results of all samples included in the flight. The performance of the proposed method is validated on the real continuous monitoring data, and the results confirm that the proposed model overcomes the problems of multistage and multivariate parameters in the anomaly detection of the aircraft system and improves the detection efficiency.

Introduction

Prognostics and health management (PHM) is an advanced technique to detect anomalies, diagnose faults, and predict the remaining useful life (RUL) for mechanical engineering systems.[1] In this emerging field, researchers have proposed concepts, theories, methods, and application frameworks related to PHM and have achieved great success on the laboratory data.[2–4] Synchronously, some novel concepts and techniques were widely used in the industry such as the intelligent health monitoring system and the novel concept of predictive maintenance.[5–7]

Considering the application of PHM methods on the aviation, the complex flight environments and the multiple monitoring parameters are the huge challenge. PHM methods, such as anomaly detection,[8] fault diagnosis,[9,10] and prognosis[11] for aircraft, have been deeply studied based on hypothesis, simulation, or laboratory data, while the modeling and learning on actual flight data are still in the exploratory stage. On one hand, in actual flight collected by the airborne quick access recorder (QAR), the proportion of normal and abnormal samples is unbalanced that why most supervised methods are not available for the aircraft monitoring and diagnosis. On the other hand, the QAR data have the characteristics of timing and high-dimensionality, and general algorithms hardly effectively handle with the time series information contained in the QAR data. In order to detect the health status of the aircraft based on the actual flight data, this article proposes an anomaly detection for the aircraft hydraulic system by the auto-encoder (AE) model with the long short-term memory (LSTM) layers, called LSTM-AE anomaly detection model. In general, the AE is a pretty

Key Laboratory of Health Monitoring and Intelligent Maintenance, Nanjing University of Aeronautics and Astronautics, Nanjing, China

Corresponding author:
Hongsheng Yan, Key Laboratory of Health Monitoring and Intelligent Maintenance, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, Jiangsu, China.
Email: rms_yhs@nuaa.edu.cn

Table 1. Summary of anomaly detection methods.

Method type	Technology used	References	Application domain	Data source
Statistical-based	Fuzzy K-means clustering + extended Kalman filter	Garg and Batra[14]	Traffic issue	Benchmark data sets
	Bayes + outlier detection	Titouna et al.[15]	Wireless sensor	54 Mica2Dot sensors data
	Gaussian mixture model + clustering	Li et al.[16]	Flight safety	Flight data
Machine learning	Local outlier probability (LoOP)	Oehling and Barry[17]	Flight safety	Monitoring data
	One-class SVM	Gharoun et al.[18]	Turbofan engine	Flight data
	Bayes + KNN classifier	Pajouh et al.[19]	Network detection	NSL-KDD data sets
Deep learning	Fully convolutional neural networks (FCNs)	Sabokrou et al.[20]	Crowded scenes	UCSD Ped2 and subway data
	Denoising auto-encoder	Jiang et al.[21]	Wind turbine	Simulation
	Convolutional denoising auto-encoder	Fu et al.[22]	Aircraft engine	ACARS data
	Deep auto-encoder	Ma et al.[9]	Actuator	Simulation

SVM: support vector machine; KNN: k-nearest neighbor.

practical unsupervised or semi-supervised algorithm that can be used to handle with unlabeled data. The LSTM is a variant of recurrent neural network (RNN) and realizes information mining of time series data by three specific gate transmission mechanisms.[12] Based on the fusion of the two advantages, the proposed LSTM-AE can effectively solve the problem of detecting the health status of the aircraft based on actual flight data.

For the aircraft, the hydraulic system is one of the most important systems in the aircraft which interacts with other aircraft systems such as the flight control system and the landing gear system. For the early failure or abnormality of the hydraulic system, it is difficult to have an intuitive display in the airborne maintenance system, and it is not easy to find it by manual inspection. The LSTM-AE model is proposed to detect early failures and anomalies of the hydraulic system based on the QAR data. Moreover, this research provides a strong illustration and feasible technical guidance for the efficient application of QAR data in engineering. Compared with some commonly used anomaly detection methods, the case study shows the efficiency and accuracy of the proposed method.

The rest of this article is organized as follows. A review of anomaly detection and related works is introduced in section "Anomaly detection and related works." In section "Proposed model," the modeling and learning processes of LSTM-AE model are described in detail. Section "The case study" shows a case to illustrate the efficiency and accuracy of the proposed detection method. Finally, the conclusive statement is in section "Conclusion."

Anomaly detection and related works

The target of anomaly detection is to find unexpected patterns in observation data based on models or data-driven approaches. These nonconforming patterns are often regarded as outliers, discordant observations,

exceptions, anomalies, aberrations, surprises, or early faults in different application domains.[13] Commonly, the anomaly is the term which is widely used in industrial processes, and anomalies present unexpected diagnostic results in this article. The result is generally determined by the similarity of label or the value of anomaly score to detect whether it belongs to the unexpected pattern.

Obviously, if the training data are labeled, the anomaly detection process is a classification process, which is a simplified process. However, the label is not always available in real scenarios, especially industrial process. Based on the availability of label, anomaly detection methods can be summarized as three types: supervised anomaly detection, semi-supervised anomaly detection and unsupervised anomaly detection. However, anomaly detection methods can be roughly divided into different groups, because of different algorithms such as statistical-based, machine learning, deep learning, and hybrid methods. A brief review of anomaly detection methods is shown in Table 1.

As described by current anomaly detection methods, the following challenges exist in the general industrial process of anomaly detection: fuzzy decision boundary, high-dimensional visualization, noise and missing data, unbalanced labels, and so on. Comparatively, the flight of an aircraft is more complex than a general industrial process, because the aircraft is affected not only by degraded mechanical components and the multistage flight process but also by the interference of external environment. Based on the existing data-driven anomaly detection methods, this article proposes an unsupervised LSTM-AE anomaly detection model for the aircraft hydraulic system, and main contributions of this research are highlighted as follows:

1. A new LSTM-AE anomaly detection model is proposed to mining temporal information and detects anomalies or early faults based on the flight QAR data.
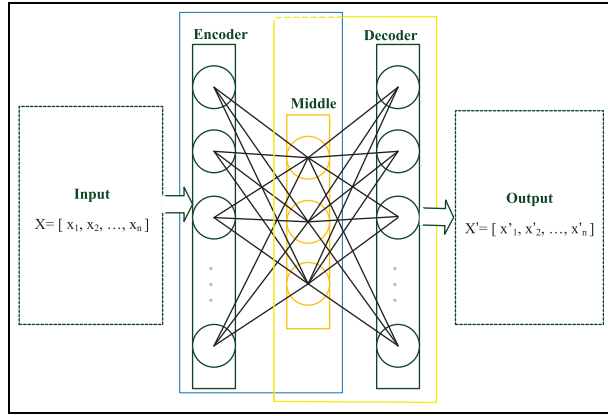
Figure 1. A simple structure diagram of auto-encoder model.

2. A novel process is used to improve the availability of QAR data based on data segmentation and recombination.
3. A case demonstrates the validity and accuracy of the proposed anomaly detection method in the real QAR data.

Proposed model

The structure of LSTM-AE mode

In order to adapt to the extreme imbalance between the proportion of normal and abnormal samples in practical engineering systems, the semi-supervised auto-encoder model for anomaly detection is widely studied and applied. The general structure of AE is a symmetrical neural network, which consists of two subnetworks: encoder and decoder. The simple structure diagram of AE model is shown in Figure 1.

In this article, considering the time continuity of QAR data within a flight, encoder and decoder, consist of LSTM layers called LSTM-AE model, are used for anomaly detection with a specific process. The LSTM is an extension of the RNN and includes three specific memory cells to acquire and transmit historical information.[23,24]

Specifically, the application mechanism of the proposed LSTM-AE anomaly detection model can be divided into three steps: the transformation of time series data, LSTM-AE model training and learning, and anomaly detection for the new observation, and the schematic diagram is shown in Figure 2. Furthermore, the detailed technical description and derivation process will be expressed in the following sections.

Model training

Data processing. Before training model, in order to meet data characteristics and model requirements, the raw QAR data should be processed: interpolation, segmentation and recombination, and normalization.

Interpolation is a statistical analysis method to supplement missing data reasonably because recording frequencies of parameters in the raw QAR data is not consistent. Typically, there are many interpolation methods such as linear interpolation, nearest neighbor interpolation, and Gauss kernel interpolation. Considering trends of different QAR parameters, appropriate interpolation methods are selected to supplement different missing values.

As described by Wang et al.,[25] the raw data are divided into several stages in a flight cycle: pre-climb, climb, cruise, descent, and ground. The purpose of the anomaly detection model proposed in this article is to discover anomalies in a flight cycle. Therefore, for this kind of multistage monitoring situations with obvious changing parameters, it is necessary to reorganize the raw data through segmented data to satisfy that each observation covers characteristics of all phases. During the pre-climb phase, the time is short and the record is incomplete, so it is not considered to add this stage
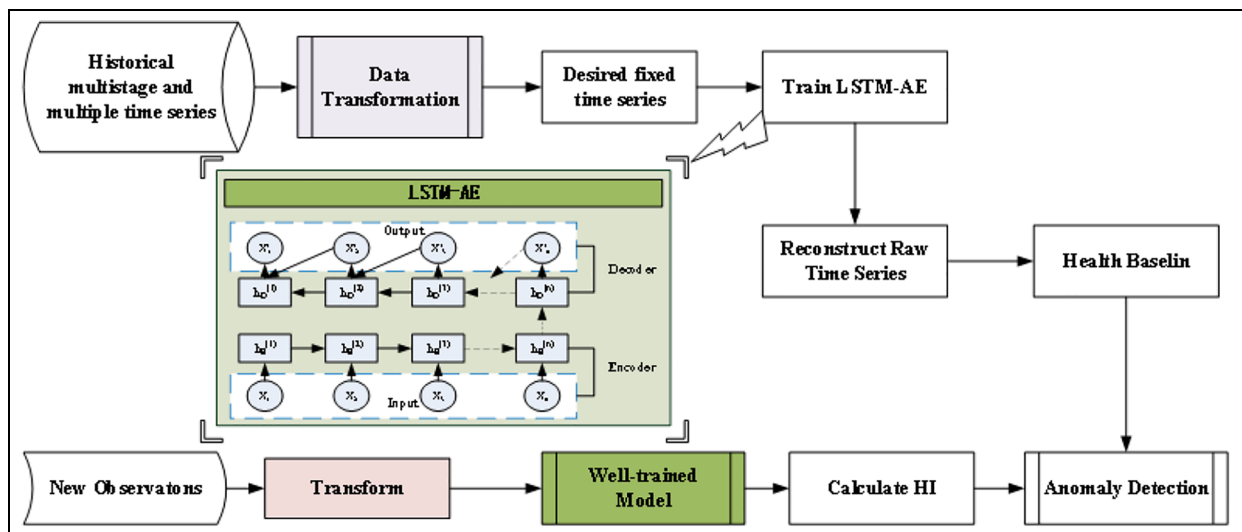

Figure 2. The schematic diagram of the application mechanism for the proposed LSTM-AE model.

when combining multistage sensor data. Given a QAR monitoring set of a flight cycle $X \in R_{n \times d}$, d represents the dimension of the raw data and n represents the length of the raw data and is divided into four parts: the length n_1 of the climb, the length n_2 of the cruise, the length n_3 of the descent, and the length n_4 of the ground. The minimum value m in the length set $\{n_1, n_2, n_3, n_4\}$ is set to the length of the reconstructed observation, and the dimension of the new observation is four times of the original data. The segmentation and reorganization of the original data are shown in equation (1)

$$X^{n \times d} = \begin{cases} X_1^{n_1 \times d} \\ X_2^{n_2 \times d} \\ X_3^{n_3 \times d} \\ X_4^{n_4 \times d} \end{cases} \xrightarrow[\substack{m = \min(\{n_1, n_2, n_3, n_4\})}]{\text{Segmentation and recombination}}$$

$$\{X_1^{m \times d} \quad X_2^{m \times d} \quad X_3^{m \times d} \quad X_4^{m \times d}\} = X^{m \times 4d} \tag{1}$$

With segmentation and recombination, the new observation of a flight cycle replaces the raw data and is used as the input to the model. For better model training, a maximum–minimum standardization method is applied to normalize input data.

The training process of the LSTM-AE model.

We train the LSTM-AE model to reconstruct fixed length sequences with minimal errors between inputs and outputs. Given an input X, the model will transform the specified shaped input X' instead of the original input X with a reshape layer and obtain the hidden representation h with a nonlinear mapping

$$h = f(W_1 \times X' + \sigma_1) \tag{2}$$

$$\widetilde{X} = g(W_2 \times h + \sigma_2) \tag{3}$$

Setting the original data set $\{Xi\}_{i=1}^{n}$, the reconstruction error of a single observation can be calculated by mean squared logarithmic error (MSLE). The training goal, which is also called the loss function of training, is to minimize the mean reconstruction error as shown in equation (4) with an optimal set of training parameters $\theta^* = \{W_1^*, \sigma_1^*, W_2^*, \sigma_2^*\}$

$$MSLE = \underset{\theta^*}{\text{minimize}}$$
$$\left(\text{mean}\left(\sum_{i=1}^{n}\left(\log(x_{\text{true}}^i + 1) - \log(x_{\text{reconstruction}}^i + 1)\right)^2\right)\right) \tag{4}$$

Learning with the LSTM-AE model.

In order to improve the robustness of the model, this article divides all data into three parts: $\{S_T, S_C, S_N\}$, in which S_T is the data set based on all health condition for training to get

parameters of the LSTM-AE model, S_C is the data set for reconstructing normal data to get the reconstruction error vector and the health threshold, and S_N is the data set with normal data and some known anomalies for testing. Specifically, although both S_T and S_C are sampled under normal conditions, S_T and S_C consisted of 70% and 20% of the health data, respectively. The remaining 10% health data and some known anomalies constitute the test set.

With the LSTM layer, the input is a three-dimensional tensor, and its shape can be defined as (N, L, K), where N is the number of samples, L is the length of each sample, and K is the number of modeling variables. Given the input $X = \{xi\}_{i=1}^{N}$, each sample is a fixed-size time series data $xi \in RL \times K$ ($i \in (1, N)$), and the reconstruction result of the LSTM-AE is $\tilde{x}i$. In this article, for each sample, residual characterized as R is used to represent the reconstruction error (refer to equation (5)), and cosine distance characterized as C is used to measure the difference between the input data xi and the reconstructed data $\tilde{x}i$ (refer to equation (6))

$$R_i = mean\left(\sum_{l=1}^{L}\sum_{k=1}^{M}\left(|\widetilde{x}(l,k) - x(l,k)|\right)\right) \tag{5}$$

$$C_i = 1 - \text{mean}$$
$$\left(\sum_{k=1}^{M}\left(\frac{\sum_{l=1}^{L} x(:,k)_l \cdot \tilde{x}(:,k)_l}{\sqrt{\sum_{l=1}^{L} x(:,k)_l^2} \cdot \sqrt{\sum_{l=1}^{L} \tilde{x}(:,k)_l^2}}\right)\right) \tag{6}$$

Next, threshold values h_R and h_C of the residual and the cosine distance should be determined for detection, respectively. In this study, a nonparametric method named kernel density estimation (KDE) is applied to directly estimate the probability density function (PDF) of the characteristic parameter. The KDE is a well-applied approach to estimate the statistical characteristics and the PDF based on the data sample itself and does not need any prior knowledge and assumptions of data distribution.[26] Given an array $\{xi\}_{i=1}^{n}$, the estimated probability density at any point x with the KDE is calculated

$$p(x) = \frac{1}{Nb}\sum_{i=1}^{N}K\left(\frac{x - x_i}{b}\right) \tag{7}$$

where $K()$ is the kernel function, and b is the bandwidth. In this study, the optimal value of b is determined with reference to previous studies—a method named mean integrated squared error (MISE), its objective is to minimize the L2 risk function.

With a given confidence level α, the optimal threshold values h_R and h_C can be calculated by the PDFs of characteristic parameters based on the following equation

$$P(x < h) = \int_{-\infty}^{h} p_b(x)dx = \alpha \qquad (8)$$

The overall scheme of anomaly detection

Totally, the anomaly detection scheme is based on the comparison of numerical sizes between the reconstructed results of unknown state observations and the health threshold which is obtained by training and analyzing the normal data.[19] Given a well-trained LSTM-AE model, the numerical performance of the reconstructed result can represent the health condition of observation with the unknown state. Unlike a single decision-making factor in previous studies, in this article, a compromise criterion function based on risk management is proposed to represent the final decision threshold function, as shown in equation (9)

$$F = k \cdot h_R + (1 - k) \cdot h_C, \quad k \in (0, 1) \qquad (9)$$

where k is the compromise coefficient, which is set to adjust the value of the function, that is, setting the best threshold function to $F^*(h_R, h_C)$ with the optimal compromise coefficient k^*.

For normal samples, because of the stability and normality of data, the reconstructed result is in good agreement with the raw data, and it will make a low reconstruction error and a high similarity between the reconstructed result and the raw data. However, the test sample with a high reconstruction error and a low similarity will be recognized as an anomaly or a fault. The process of the anomaly detection based on the LSTM-AE is shown in Figure 3.

The above introduction is aimed at the detection for individual input sample, but hardly meets the detection requirement of the whole flight QAR data that contains several samples. The normal sample is marked as "0," the abnormal sample is marked as "1." In this article, we propose the health index (HI) based on voting regulation to present the normal proportion of one flight. When a flight contains more than the specified percentage of abnormal samples, HI is below the reliability level and the flight is considered as unhealthy. However, if the flight is considered as healthy, HI can be presented as HI = $I(0)/(I(0) + I(1))$.

Performance metric

In general, for the results of anomaly detection, true positive (TP) presents the number of abnormal cases correctly detected as anomalies, while true negative (TN) presents the number of normal cases correctly detected as normal. At the same time, false positive (FP) denotes the number of normal cases misdiagnosed as abnormal, and false negative (FN) denotes the number of abnormal cases misdiagnosed as normal. To evaluate the performance of the anomaly detection method, some metrics have been proposed and widely used such as accuracy (the ratio of correctly diagnosed cases to the total observations), precision (the ratio of correctly diagnosed positive cases to the total
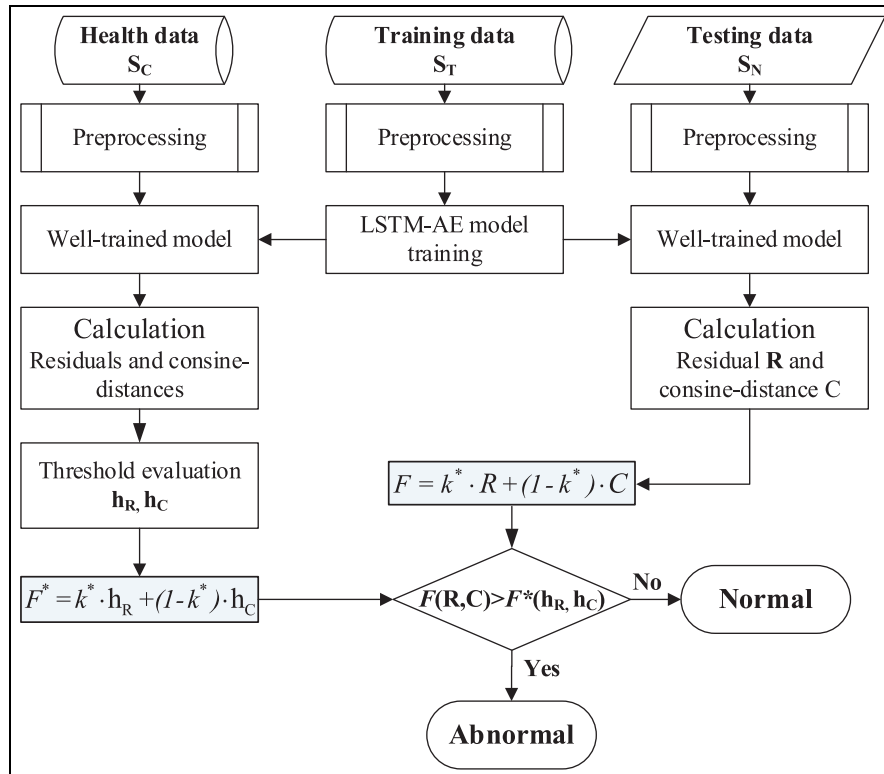

Figure 3. The process of anomaly detection based on LSTM-AE.

diagnosed positive observations), and recall rate (the ratio of correctly predicted positive cases to the total actual positive observations). These metrics can be computed by the following formulas

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (10)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (11)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (12)$$

Although the accuracy can determine the overall accuracy rate, it cannot be used as a good indicator to measure the results when the sample is not balanced. The precision is for the detection results, meanwhile, the recall focuses on the original observation, and they have different concerns. Usually, the relationship between the precision and recall is usually negatively related. Absolutely, the proportion of normal and abnormal samples is unbalanced in the data set, and an effective and applicable comprehensive indicator is very important for evaluating the model. Consequently, F-score is selected as the evaluation indicator, which is the weighted fusion value of the precision and the recall,[14] and the F-score can be calculated by equation (13)

$$\text{F} - \text{score} = (1 + \beta^2) \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (13)$$

where β is the balance coefficient, which presents the importance between the precision and the recall. When $\beta = 1$, it means that the precision is as important as the recall. When $\beta < 1$, the precision is more important than the recall. Conversely, when $\beta > 1$, the recall is more important than the precision. In this study, for anomaly detection, the precision is as important as the recall, that is, the value of β is 1.

The case study

In this article, we consider real flight data of the hydraulic system, the Boeing's B737 aircraft, to validate the effectiveness and accuracy of the proposed LSTM-AE anomaly detection method.

Introduction of the hydraulic system

The single-aisle B737 aircraft is one of the most popular commercial aircrafts, and it has a set of mature and critical hydraulic system, which covers three subsystems: the main hydraulic subsystems A and B (the two main subsystems have same components and linked by the specific device) and the auxiliary hydraulic subsystem. In general, the auxiliary hydraulic subsystem only works and provides the necessary power for related hydraulic power users when both the main hydraulic subsystems failed to work normally. In fact, there are

no parameters related to the auxiliary hydraulic subsystem in the QAR data of the B737.

The B737's hydraulic system includes oil reservoir and its accessories, power pump (engine drive pump (EDP) and electric motor drive pump (EMDP)), pipeline, filter modules, heat exchanger, power transfer unit (PTU, one-way, this module can only transfer the pressure from A to B when subsystem B cannot provide hydraulic energy normally), pressure module, and valves. Take the main subsystem A of the B737NG aircraft as an example, refer to the aircraft maintenance manual (AMM) of this aircraft type, the main components and structure of the hydraulic system are shown in Figure 4.

There are many kinds of fault modes and causes of the aircraft hydraulic system. In engineering practice, the main faults of the hydraulic system are abnormal pressure output, internal and external oil leakage, and hydraulic oil pollution (manifested as filter blockage).[27] Common faults of the hydraulic system are shown in Table 2.

Considering the consequences of faults and the difficulty of identification, the fault of oil leakage is most time-consuming for fault isolation and easily leads to other more serious failures. The general process of hydraulic oil leakage is as follows: oil percolation in a small internal scope (step 1)—droplet leakage in a limited scope (step 2)—serious leakage in a large scope (step 3).[28] The purpose of this article is to construct an anomaly detection model based on multivariate monitoring data, which can be used to identify anomalies in the first or second step of the degradation process of hydraulic system leakage fault and avoid serious leakage fault.

Data and model structure description

The real QAR data of the aircraft hydraulic system are used to train and test the LSTM-AE model. The definition and description of the parameters of the hydraulic system are shown in Table 3.

One important issue is to select available and efficient interpolation methods for different numerical parameters, such as HOQA, HOQB, PASA, and PRSB, because the record frequency of these parameters is not once per second. Its purpose is to improve the quality of the model input data while ensuring the efficiency of data processing and to meet the data trend and the engineering experience of airline maintenance. In general, simple and practical linear interpolation and nearest neighbor interpolation methods are widely used in the processing of industrial data, the latter is more applied to image data. For HOQA and HOQB, the record frequency is once per 4 s, and the trend is linear and continuous in the local window, so the linear interpolation method is selected to process these two parameters. For PRSA and PRSB, as the pressure parameter of subsystem, the record frequency is once per 2 s, and the trend is nonlinear and fluctuating.
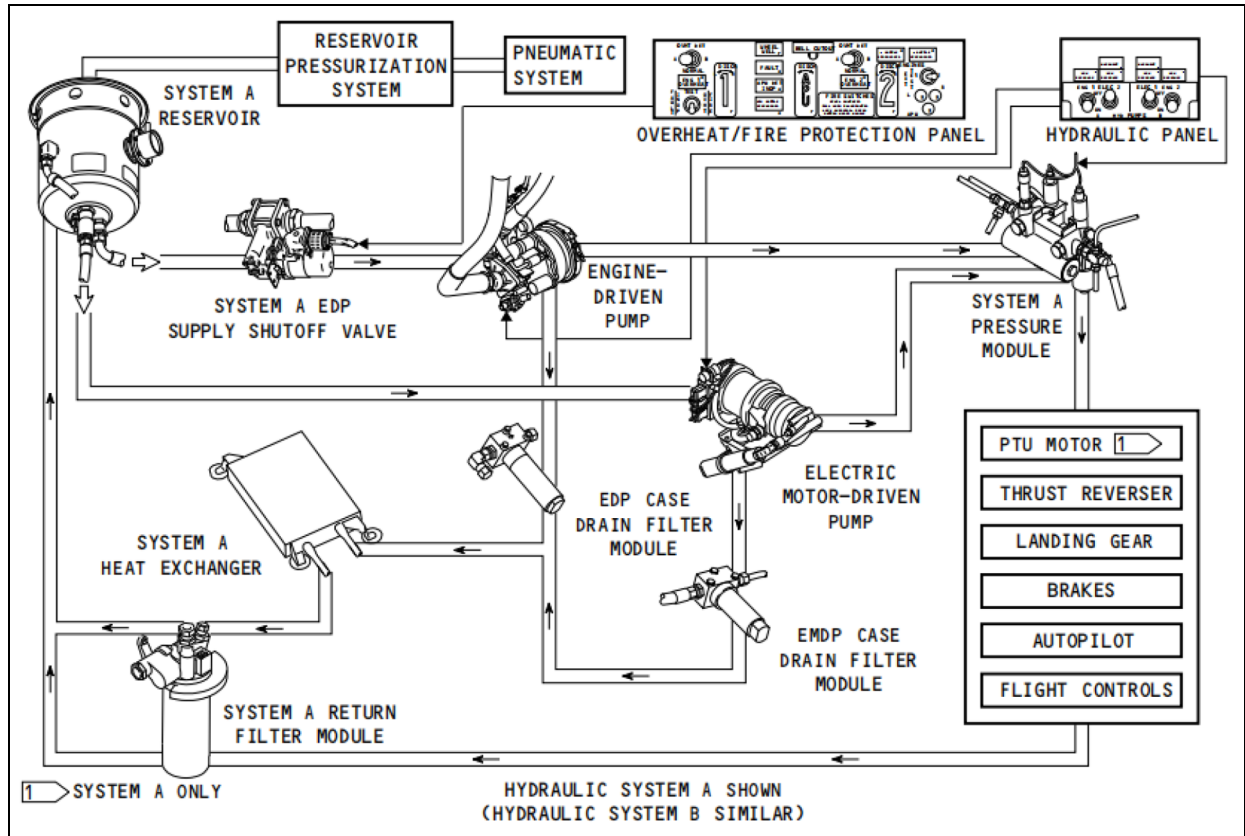
Figure 4. The schematic diagram of the main hydraulic subsystem A.

Table 2. Summaries of common faults of the hydraulic system..

Fault description	Effect	Fault cause analysis
System low-pressure alarm	The hydraulic system users lose the necessary pressure source and face a hard operation in some serious cases	Stall of hydraulic pumps and functional failure of the PTU
Plugging of the oil filter	Damages filter units and affects the hydraulic pressure output	Functional failure of filters and hydraulic oil pollution
Hydraulic oil leakage	Affects the hydraulic pressure, pollution, and damage- related units	Leakage of the oil tank, leakage of pipelines, and leakage of hydraulic actuators

PTU: power transfer unit.

Table 3. Definition and description of parameters of the hydraulic system.

Parameters	Description	Unit	Record frequency	Interpolation method
N11	Low-pressure rotor speed of left engine	r/min	Once per second	None
N12	Low-pressure rotor speed of right engine	r/min	Once per second	None
N21	High-pressure rotor speed of left engine	r/min	Once per second	None
N22	High-pressure rotor speed of right engine	r/min	Once per second	None
HOQA	Oil quantity of system A	L	Once per 4 s	Linear interpolation
HOQB	Oil quantity of system B	L	Once per 4 s	Linear interpolation
PRSA	Pressure of system A	psi	Once per 2 s	Nearest neighbor interpolation
PRSB	Pressure of system B	psi	Once per 2 s	Nearest neighbor interpolation
HSA1	EDP low-pressure of A	None	Once per second	None
HSAE	EMP low-pressure of A	None	Once per second	None
HSB2	EDP low-pressure of B	None	Once per second	None
HSBE	EMP low-pressure of B	None	Once per second	None

EDP: engine drive pump; EMP: electromagnetic pulse.

Table 4. Layer structure of the proposed LSTM-AE.

Layer	Output shape	Description
input_1	(None, 50, 32)	Input layer
lstm_1	(None, 50, 100)	LSTM encoding layer 1
dropout_1	(None, 50, 100)	Prevent overfitting
lstm_2	(None, 50, 50)	LSTM encoding layer 2
dropout_2	(None, 50, 50)	Prevent overfitting
lstm_3	(None, 50, 10)	Embedding layer
lstm_4	(None, 50, 50)	LSTM decoding layer 1
dropout_3	(None, 50, 50)	Prevent overfitting
lstm_5	(None, 50, 100)	LSTM decoding layer 2
dropout_4	(None, 50, 100)	Prevent overfitting
lstm_6	(None, 50, 32)	Output layer

LSTM: long short-term memory; AE: auto-encoder.

Referring to the processing method for image data, the simple and feasible nearest neighbor interpolation is used to deal with missing values of pressure parameters.

In general, these low-pressure warning parameters are all "normal" in the QAR record, and only when the major failures occur, such as the description of serious leakage in the last section, the alarm "low-pressure" will be reported. The proposed LSTM-AE model is used to identify early and limited leakage faults, which are unresponsive in the alarm parameters. For modeling and training, the eight parameters (N11, N12, N21, N22, HOQA, HOQB, PRSA, and PRSB) in the QAR data of the hydraulic system are selected as initial characteristic parameters, and the model input data with fixed shape have been obtained by data transformation.

As the description of data processing in model training, considering the input parameters are eight types and including four flight stages, the fixed input shape is the 32-dimensional form. The network layer structure of LSTM-AE in this article is shown in Table 4.

The LSTM-AE model is compiled and configured in Keras by Python, and the hardware condition is a regular computer with specification parameters as i7-855U of CPU, 8 GB of RAM, and 64-bit operating system of the windows-10. Adam is selected as the optimizer and MSLE is selected as the loss function. Based on the set network structure, the training error of the LSTM-AE model training process is shown in Figure 5. The training error tolerance is set to 0.01, so the training process will be stopped when the training error goes below 0.01.

Results and analysis

With the well-trained model, we can get the residual list and the cosine distance list of reconstructed results based on the health data S_C. Then, the threshold values h_R and h_C of the residual list and the cosine distance list can be calculated by the KDE, respectively, and the Gaussian kernel is selected as the kernel function as shown in equation (14)

$$K(x) = \frac{1}{\sqrt{2\pi}} exp\left(-\frac{x^2}{2}\right) \tag{14}$$
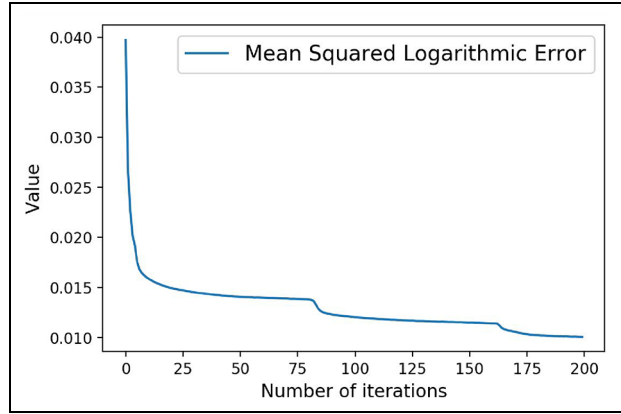
Figure 5. The training error of the LSTM-AE.
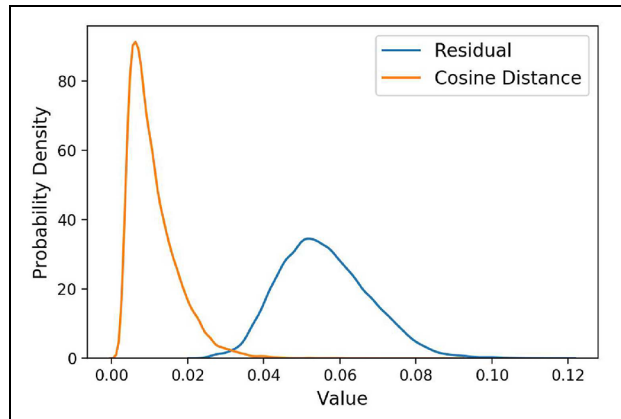
Figure 6. The probability density of residual and cosine distance of the health data.

Satisfying the minimum of MISE, the optimal bandwidth of the KDE can be obtained, and the optimal bandwidths of the residual and the cosine distance are 0.0012 and 0.0037, respectively. The probability densities of the residual and the cosine distance by the KDE are shown in Figure 6.

Given the confidence level to 95%, the threshold value of the residual and the cosine distance can be obtained based on the well-fitted cumulative probability curve, as shown in Figure 7.

As shown in Figure 7, the single threshold values of the residual and the cosine distance are 0.076 and 0.023 with the 95% confidence level, respectively. Then, the decision threshold function $F_k(h_R, h_C)$ is determined by the single threshold value h_R and h_C based on equation (9) with a compromise coefficient k. Furthermore, the best decision threshold, which fuses both hR and h_C, is used to measure the health status of a new sample.

As the rule of anomaly detection shown in section "The overall scheme of anomaly detection," the new observation is defined as an anomaly when its reconstructed result goes beyond the threshold function; on the contrary, one is regarded as a normal condition when the result goes below the threshold.
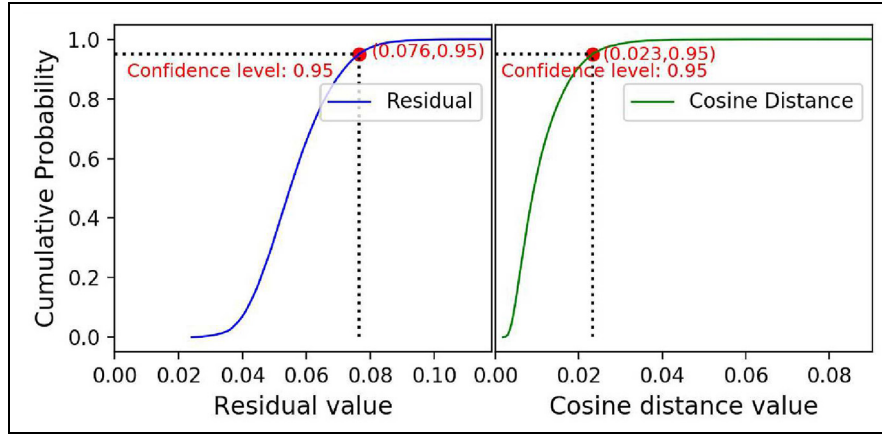
Figure 7. The cumulative probability curve of the residual and the cosine distance of the health data.

Table 5. Comparing model performance under different detection indicators.

Detection indicators		Accuracy	Precision	Recall	F-score
Single threshold	Residual (k = 1)	0.928	0.856	0.884	0.870
	Cosine distance (k = 0)	0.929	**0.915**	0.813	0.861
Comprehensive threshold (optimal k with maximum F-score value)		**0.958**	0.912	**0.934**	**0.923**

The bold values mean the maximum value in this category.

With the compromise coefficient k (in the range of 0–1), we use the comprehensive threshold function to distinguish anomalies. In order to obtain the optimal k, we use the repetitive numerical simulation with fixed step to calculate the model performance for anomaly detection based on the test data set S_N (includes 10,537 samples). From the simulation result shown in Figure 8, each curve represents a performance metrics of the anomaly detection model based on the data set S_N. The model performances under different detection indicators are shown in Table 5.

Obviously, the detection performance of a single detection indicator is not as good as that of the comprehensive threshold function. Only using the residual as detection indicator, it makes a lower precision which is related to the missing alarm rate of anomaly detection (the lower precision and the higher missing alarm rate). Similarly, it makes a lower recall with only the cosine distance, which leads to a high false alarm rate of anomaly detection. In this article, the performance metrics F-score, which is calculated by equation (13) with precision and recall, is used to determine the optimal coefficient k.

Validation of the real continuous monitoring data

With the data processing, the real continuous monitoring data can be transformed into samples in the required format of the model, and the healthy degree of these samples can be defined according to the after-flight inspection and maintenance report. Taking oil quantity of the subsystems A and B (HOQA and
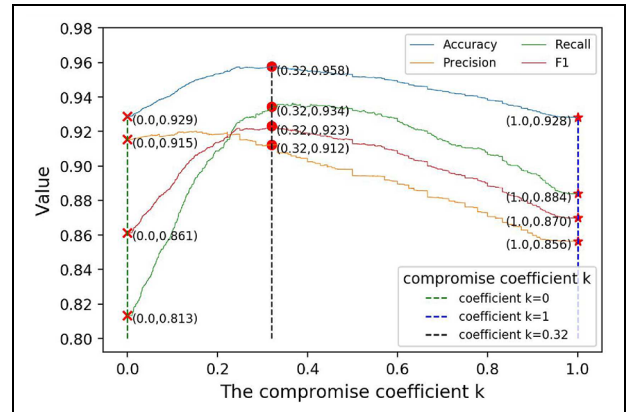

Figure 8. Model performance for anomaly detection with different coefficient k.

HOQB) as examples, Figure 9 shows the dynamic change in parameters.

Each flight is divided into several input samples, and the reconstructed results of each sample can be obtained through LSTM-AE model. According to the anomaly detection rule, the HI of each flight can be calculated. The detection results of these flights are shown in Figure 10.

In this study, the actual record of fault inspection and maintenance process that occurred in the aircraft hydraulic system is shown in Table 6.

In this continuous QAR monitoring data, the daily record contains several flights, each of which is regarded as an observation sample. According to the
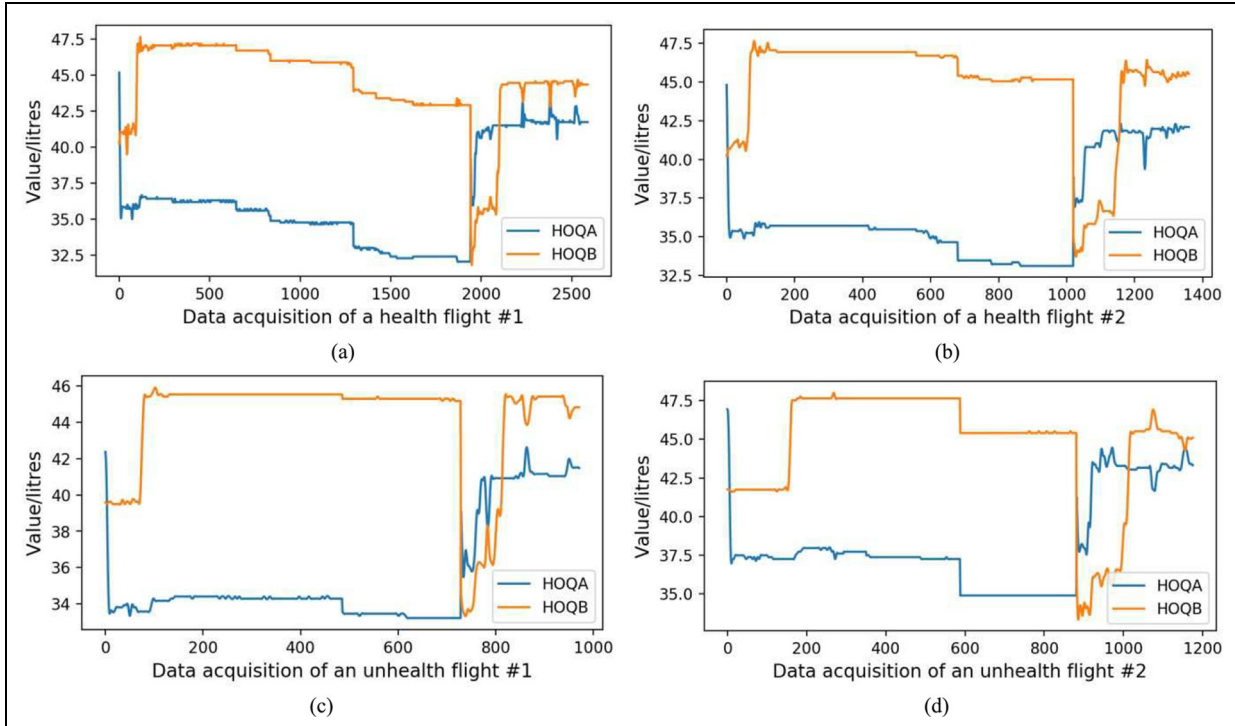
Figure 9. Data acquisition of the hydraulic subsystem oil quantity ((a) and (b) show dynamic changes in monitoring parameters in the normal flight, while (c) and (d) show dynamic changes in monitoring parameters in the abnormal flight).
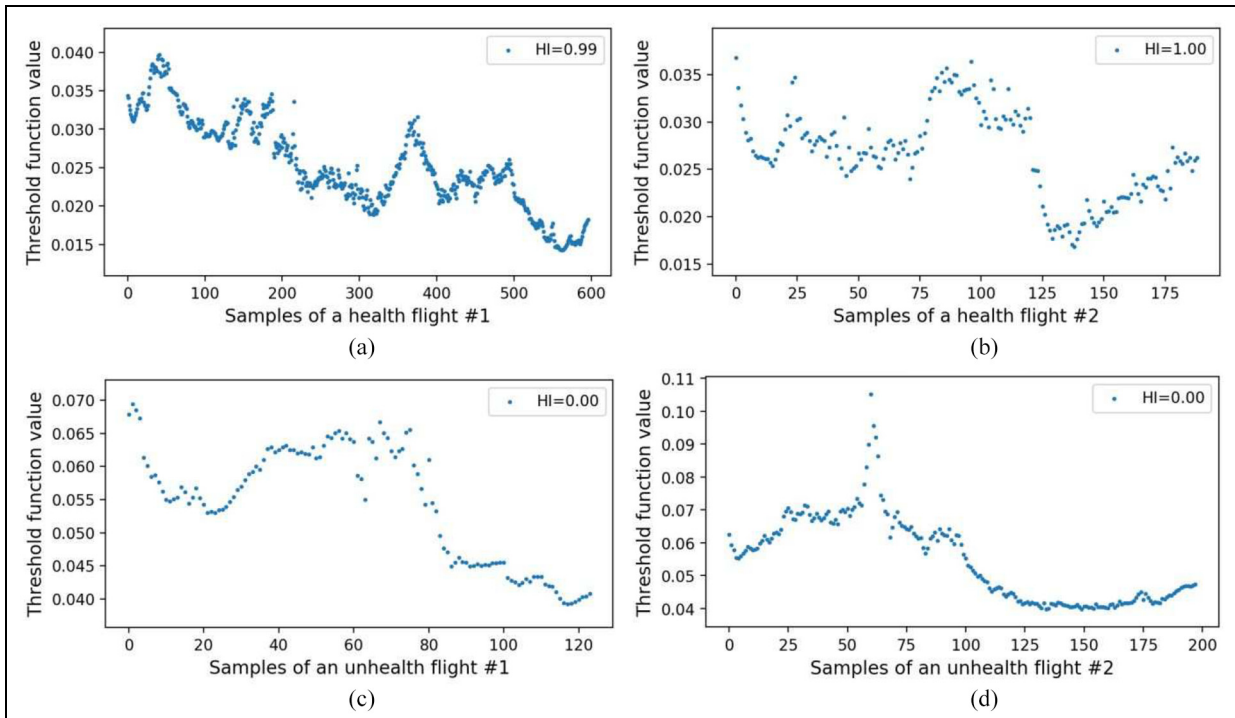

Figure 10. The health indexes of instance flights: (a and b) show health indexes of health flights, and (c and d) show health indexes of unhealth flights.

data processing method introduced in this article, the original record is converted into model input, and the comprehensive function value of model output is shown in Figure 11.

Flight status is defined as abnormal when the value of a reconstructed comprehensive function exceeds the threshold. As shown in Figure 11, there are two points of the reconstructed results of test set by the LSTM-AE

Table 6. Fault detection and maintenance process records of the hydraulic system.

Record number	Inspection date	Fault description	Maintenance process	Handling date
1	18 Aug 2016 00:00:00			18 Aug 2016 01:10:00
		Oil leakage of the outlet plug of the hydraulic system A electric pump	Replace the seal ring of the system A EMDP	
2	23 Oct 2016 13:12:37			23 Oct 2016 14:10:00
		Oil leakage of the outlet plug of the hydraulic system A electric pump	Replace the seal ring of the system A EMDP	

EMDP: electric motor drive pump.
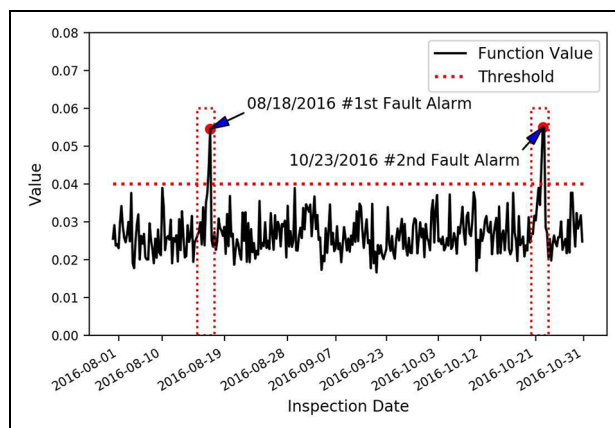
Figure 11. The comprehensive function value of detection of the continuous data.

model exceed the threshold, near 18 August 2016 and 23 October 2016, respectively. Obviously, the value of reconstructed comprehensive function increases continuously in a few flights before the anomaly, and because of timely inspection and maintenance, the value after the anomaly quickly returns to stable fluctuation. Referring to the fault inspection records in Table 6, the detection result based on the LSTM-AE model proposed in this study is perfectly close to the actual situation. It shows that the anomaly detection method proposed in this article has a good early warning effect for fault detection and can guide the maintenance schedule.

Conclusion

In general, the main contributions of this article are concluded as follows: (1) For the better performance of the anomaly detection model, this article proposed the LSTM-AE model for mining historical information of time series data and realizing an unsupervised learning process without available labels. The LSTM-AE provides a new perspective to improve the role of the flight data in the aviation industrial practice, such as flight health detection, failure diagnosis, and RUL prediction, based on fusion methods; (2) For improving the availability of multistage and high-dimensional flight

QAR data, a novel processing method based on data segmentation and recombination is proposed. Through two steps of segmentation and recombination, the original data are effectively redefined to the new input data; (3) A comprehensive decision-making index which fuses two decision variables, such as residual and cosine distance, is proposed, and the threshold is obtained by the KDE.

Finally, based on the conclusion of this article, some future research works are discussed. In the first step, based on further excavation and analysis of the output of the hidden layer, combined with the current usage time, the research of predicting RUL will be carried out. Then, considering the results of anomaly detection and RUL prediction, combined with actual maintenance parameters, such as initial inspection interval, single inspection cost, replacement component cost, inventory cost, with availability and total cost as the goal, the predictive maintenance plan and interval optimization research will be carried out.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ORCID iDs

Hongsheng Yan   https://orcid.org/0000-0001-8675-6121

Jianzhong Sun 🆔 https://orcid.org/0000-0003-1806-7388

References

1. Lee J, Wu F, Zhao W, et al. Prognostics and health management design for rotary machinery systems—reviews, methodology and applications. *Mech Syst Signal Pr* 2014; 42(1–2): 314–334.
2. Compare M, Bellani L and Zio E. Reliability model of a component equipped with PHM capabilities. *Reliab Eng Syst Safe* 2017; 168: 4–11.
3. Wang Z, Cui Y and Shi J. A framework of discrete-event simulation modeling for prognostics and health management (PHM) in airline industry. *IEEE Syst J* 2017; 11(4): 2227–2238.
4. Liu X, Li J and Al-Khalifa KN. Condition-based maintenance for continuously monitored degrading systems with multiple failure modes. *IIE Trans* 2013; 45(4): 422–435.
5. Sandborn PA and Wilkinson C. A maintenance planning and business case development model for the application of prognostics and health management (PHM) to electronic systems. *Microelectron Reliab* 2007; 47(12): 1889–1901.
6. Hongsheng Y, Hongfu Z, Jianzhong S, et al. Cost effectiveness evaluation model for civil aircraft maintenance based on prognostics and health management. In: *Proceedings of the 2017 international conference on sensing, diagnostics, prognostics, and control (SDPC)*, Shanghai, China, 16–18 August 2017, pp.161–167. New York: IEEE.
7. Jia X, Zhao M, Di Y, et al. Assessment of data suitability for machine prognosis using maximum mean discrepancy. *IEEE T Ind Electron* 2018; 65(7): 5872–5881.
8. Höhndorf L, Siegel J and Sembiring J. Reconstruction of aircraft states during landing based on quick access recorder data. *J Guid Control Dynam* 2017; 40(9): 2393–2398.
9. Ma J, Ni S, Xie W, et al. Deep auto-encoder observer multiple-model fast aircraft actuator fault diagnosis algorithm. *Int J Control Autom* 2017; 15(4): 1641–1650.
10. Liu C, Sun J, Wang F, et al. Bayesian network method for fault diagnosis of civil aircraft environment control system. Proc IMechE, Part I: J Systems and Control Engineering 2020; 234(5): 662–674.
11. Lu F, Wu J and Huang J. Aircraft engine degradation prognostics based on logistic regression and novel OS-ELM algorithm. *Aerosp Sci Technol* 2019; 84: 661–671.
12. Lipton ZC, Berkowitz J and Elkan C. A critical review of recurrent neural networks for sequence learning, 2015, https://arxiv.org/pdf/1506.00019.pdf
13. Chandola V, Banerjee A and Kumar V. Anomaly detection: a survey. *ACM Comput Surv* 2009; 41(3): 1–58.
14. Garg S and Batra S. A novel ensembled technique for anomaly detection. *Int J Commun Syst* 2017; 30(11): e3248.
15. Titouna C, Aliouat M and Gueroui M. Outlier detection approach using Bayes classifiers in wireless sensor networks. *Wireless Pers Commun* 2015; 85(3): 1009–1023.
16. Li L, Hansman RJ, Palacios R, et al. Anomaly detection via a Gaussian mixture model for flight operation and safety monitoring. Transport Res C: Emer 2016; 64: 45–57.
17. Oehling J and Barry DJ. Using machine learning methods in airline flight data monitoring to generate new operational safety knowledge from existing data. *Safety Sci* 2019; 114: 89–104.
18. Gharoun H, Keramati A and Nasiri MM. An integrated approach for aircraft turbofan engine fault detection based on data mining techniques. *Expert Syst* 2019; 36(2): e12370.
19. Pajouh HH, Dastghaibyfard G and Hashemi S. Two-tier network anomaly detection model: a machine learning approach. *J Intell Inf Syst* 2017; 48(1): 61–74.
20. Sabokrou M, Fayyaz M and Fathy M. Deep-anomaly: fully convolutional neural network for fast anomaly detection in crowded scenes. *Comput Vis Image Und* 2018; 172: 88–97.
21. Jiang G, Xie P, He H, et al. Wind turbine fault detection using a denoising autoencoder with temporal information. *IEEE/ASME T Mech* 2018; 23(1): 89–100.
22. Fu X, Luo H and Zhong S. Aircraft engine fault detection based on grouped convolutional denoising autoencoders. *Chinese J Aeronaut* 2019; 32(2): 296–307.
23. Wu Y, Yuan M, Dong S, et al. Remaining useful life estimation of engineered systems using vanilla LSTM neural networks. *Neurocomputing* 2018; 275: 167–179.
24. Habler E and Shabtai A. Using LSTM encoder-decoder algorithm for detecting anomalous ADS-B messages. *Comput Secur* 2018; 78: 155–173.
25. Wang L, Wu C and Sun R. An analysis of flight quick access recorder (QAR) data and its applications in preventing landing incidents. *Reliab Eng Syst Safe* 2014; 127: 86–96.
26. Odiowei PEP and Cao Y. Nonlinear dynamic process monitoring using canonical variate analysis and kernel density estimations. *IEEE T Ind Inform* 2010; 6(1): 36–45.
27. Liu H, Zhang J and Lu C. Performance degradation prediction for a hydraulic servo system based on Elman network observer and GMM–SVR. *Appl Math Model* 2015; 39(19): 5882–5895.
28. Zhao X, Zhang S and Zhou C. Experimental study of hydraulic cylinder leakage and fault feature extraction based on wavelet packet analysis. *Comput Fluids* 2015; 106: 33–40.

Appendix I

Notation

b	bandwidth of kernel density estimation
$CosD$	cosine distance
d	dimension of a flight raw data
f, g	activation functions
F	compromise criterion function based on risk
h	output of the hidden layer
h_C	threshold of cosine distance
h_R	threshold of residual
i	sample index
I	count of one label
k	compromise coefficient
K	kernel function
L	the length of a sample

m	minimum length of different phases in a flight	W	weight matrix
		x	fixed-size time series sample
M	variable number for modeling	X	raw data of a flight
n	length of a flight raw data	X'	specified shaped input of a flight
N	the number of samples in a flight	$\widetilde{X}$	reconstructed output
r	residual		
S_C	health data for calculating thresholds	α	confidence level
S_N	test data	σ	bias vector
S_T	train data		