Journal of Computers Vol. 34 No. 5, October 2023, pp. 29-43
doi: 10.53106/199115992023103405003

Research on Intelligent Operation and Maintenance (O&M) Method of Complex Products Based on Digital Twin

Chuanwei Zhang, Lingling Dong*, Yunrui Wang

School of Mechanical Engineering, Xi'an University of Science and Technology, Xi'an, 430000, China
dongll5932@163.com

Received 14 September 2022; Revised 6 January 2023; Accepted 10 February 2023

Abstract. With the development of advanced data analysis technology, the manufacturing industry is moving towards the direction of intelligence. To solve the problems of low operation and maintenance (O&M) efficiency, passive O&M personnel and low intelligence of complex products, a digital twin-based intelligent O&M method for complex products is proposed. A digital twin intelligent O&M model with the physical O&M center, virtual O&M center, twin data platform and O&M service system is established. In the case of product fault classification, the maintenance mechanism is designed and the O&M process in different states is analyzed. Then the implementation process and key technologies are described in detail. Through the digital twin model, the virtual-real interaction and data dynamic update of O&M are realized. Finally, taking the key components of a certain type of EMU bogie as an example, the K-means clustering and Apriori algorithm are used to analyze the fault data. Moreover, the validity and feasibility of the proposed model are verified by applying the fault data to the digital twin architecture. The proposed model and key technologies can provide a new solution for the intelligent O&M of complex products.

Keywords: digital twin, intelligent O&M, complex products, fault analysis

1 Introduction

Complex products have the characteristics of high multidisciplinary coupling, complex parts and high technical difficulty [1-3]. The operation uncertainty of complex products, the complexity of products, and the dynamic nature of maintenance resources will bring challenges to O&M management decisions. Great attention has been paid by the manufacturing industry for the improvement and optimization of product lifecycle management. As the end of the product lifecycle, O&M is the focus of customers. The intelligence level of O&M and the ability to interact with the business directly affect the lifetime of products [4]. Hence, it is urgent to transform the O&M processes of complex products into digital, informative and intelligent process.

Many researches have been carried out to improve the O&M of products. Liu et al. combined neural networks and long and short-term memory networks for fault prediction, deep reinforcement learning for production control and maintenance scheduling, and the AR technology to guide personnel through maintenance experience in a visual format [5]. P. Yepez et al. developed a knowledge-based system to improve maintenance efficiency by integrating information in FMEA into the disassembly sequences to generate specific maintenance plans, thereby guiding user maintenance [6]. K. Antoszet al. applied decision trees and rough set theory to lean maintenance of manufacturing, so as to improve the operational efficiency of a company's infrastructure through intelligent systems that support O&M decisions [7]. Mohan et al. proposed an adaptive ARIMA based on machine learning, which supports adaptive error prediction models through variable window technology. Zero downtime in industry was achieved by predicting important characteristic parameters of machinery and notifications before failures [8]. Based on the KFCM-F algorithm and the kernelised clustering validity index, Li et al. proposed a two-stage clustering framework. The simulated experiment proved that the framework can effectively discover the potential relationships between single faults, improving the efficiency in fault diagnosis [9]. Edson et al. integrated a probabilistic and predictive model constructed from time log information. Data mining techniques were used to estimate process cycle variations to optimize the maintenance inspection intervals [10]. All these studies are based on the historical fault data and the predictive models to predict the failures for preventive maintenance. The above methods are data-driven for efficient product maintenance, and they do not allow real-time monitoring of the maintenance process and the status of maintenance resources. Although the historical data is analyzed and

* Corresponding Author

predicted, the predictions may not match the real failure. In addition, rapid O&M planning and scheduling cannot be guaranteed for new faults.

With the development of communication and artificial intelligence technologies, new ideas, models, technologies and applications have been brought to the manufacturing industry [11]. To seize the commanding heights of manufacturing intelligence first, many countries have proposed different strategies. Germany proposed Industry 4.0 to achieve intelligence throughout the product lifecycle through the integration of the physical and virtual worlds [12]. The US has proposed the “Industrial Internet”, which connects people, equipment and data to form an industrialized network [13]. China has proposed Made in China 2025, which centers on smart manufacturing, green manufacturing, and whole life cycle networking and intelligence [14]. However, how to improve the collaborative business linkage capability of complex products and achieve intelligent O&M through the interaction and integration of virtual world and real world is always a matter of concern. To this end, the emergence of digital twin technology would provide a new way to solve these problems.

The concept of digital twin was originally named “information mirror model” by Michael Grieves, and later evolved into the term “digital twin” [15]. Since then, many different definitions of the digital twin have been proposed. Rosen et al. defined it as a realistic model that reflects current state processes in real world and the interaction of its own behavior in the environment [16]. Betty et al. defined it as “a mirror image of a physical process connected with the process in question”, which is usually matched exactly with the actual physical process [17]. With the introduction and development of digital twin, many scholars have gradually applied it to the whole lifecycle management of products. Tao et al. conducted a preliminary study on the application methods and framework for digital twin-driven product design, manufacturing, and services the three case studies are respectively used to illustrate possible models for future digital twin-driven product design, manufacturing, and operation and maintenance services [18]. Tao et al. proposed a five-dimensional model of digital twin technology and applied it to shop floor manufacturing [19]. Zhuang et al. proposed a digital twin-based approach for complex product assembly data management and process traceability. A digital twin-based assembly process management and control system (DT-APMCS) was designed and applied to some aerospace-related assembly companies [20]. O. Davies et al. integrated the virtual and degradation models of key components to understand the degradation status of components through digital twin technology, proposing appropriate maintenance interventions [21]. Li et al. proposed a digital twin framework for the analysis of products to be designed based on the operational data. The digital twin data is processed and analyzed by clustering algorithms, frequent growth algorithms and multi-attribute decision-making to obtain product failure data for next-generation product design [22]. Wang et al. proposed a predictive maintenance model for electromechanical equipment by introducing digital twin technology. According to setting multiple sensors in the device, the physical and spatial signals of the device are collected. The corresponding digital twin model was established based on the sensor data and a comprehensive fault diagnosis method is designed through migration learning [23]. A. Coraddu et al. established a data-driven digital twin ship. The operational data of the ship was collected by the arrangement of the corresponding sensors, and the maintenance efficiency of the ship was improved [24]. As shown in above, the digital twin technology can reflect the real-time status of product O&M by using the interaction between physical entities and virtual models of products. It can monitor the O&M status of products in real-time, predict the faults with twin data, and develop accurate maintenance plans according to the knowledge base to ensure accurate implementation of the plan and improve the quality of product maintenance. In this way, the digital twin technology can realize intelligent O&M of complex products.

Based on the application of digital twin technology in manufacturing industry, an intelligent O&M scheme for complex products based on digital twin is proposed to realize intelligent control of O&M process of complex products. The rest of this paper is structured as follows: Section 2 proposes an intelligent O&M model for complex products based on digital twins. The components and corresponding roles of the digital twin model are introduced, including the O&M entity layer, the O&M virtual layer, the twin data platform and the O&M service system. Section 3 describes the operational mechanism of O&M of complex products based on intelligent twin data, and introduces the interoperability in the process of digital twin implementation. In section 4, taking the O&M of a certain type of EMU as an example, the fault prediction of EMU bogie is analyzed by using K-means algorithm and Apriori algorithm. Combined with the data analysis, the proposed model is applicable. The feasibility and effectiveness of intelligent O&M framework based on the digital twin proposed in this paper are verified, and Section 5 gives the conclusion.

2 A Complex Product O&M Model Based on the Digital Twin

The O&M model of complex products driven by digital twin is to realize the remote O&M of products. This technology builds an ultra-high-fidelity model by embedding sensors in the product to provide a more realistic simulation of the O&M process. The real-time data of product operation collected by the sensors and the historical data of the same series of different products are transferred to the twin data platform. At the same time, the virtual O&M layer constructs the corresponding digital model of the product through the data transfer platform, which includes the physical model of the product and the environment model. The twin data platform delivers high-quality data to the O&M service system through data fusion and processing. The O&M service system uses algorithms and mathematical models for mining. It will analyze the O&M data, predict the product status, formulate the corresponding preventive maintenance plan and feed them back to the O&M resource scheduling system. The O&M resource scheduling system transfers the resource allocation information to the virtual O&M workshop to simulate the O&M process. In this process, the problems in O&M are found, and the resource scheduling plan is adjusted in time. The "combination of reality and reality" of the digital twin ensures the data integrity to be analyzed. The interaction between data, data processing, and analysis can predict the trend of product operation status, which provides a guarantee for early detection and avoidance of product operation failures. Here a complex product O&M model based on a digital twin is proposed, as shown in Fig. 1. The model mainly consists of a physical entity O&M layer, a virtual O&M model layer, an O&M twin data platform layer, and an O&M service system.

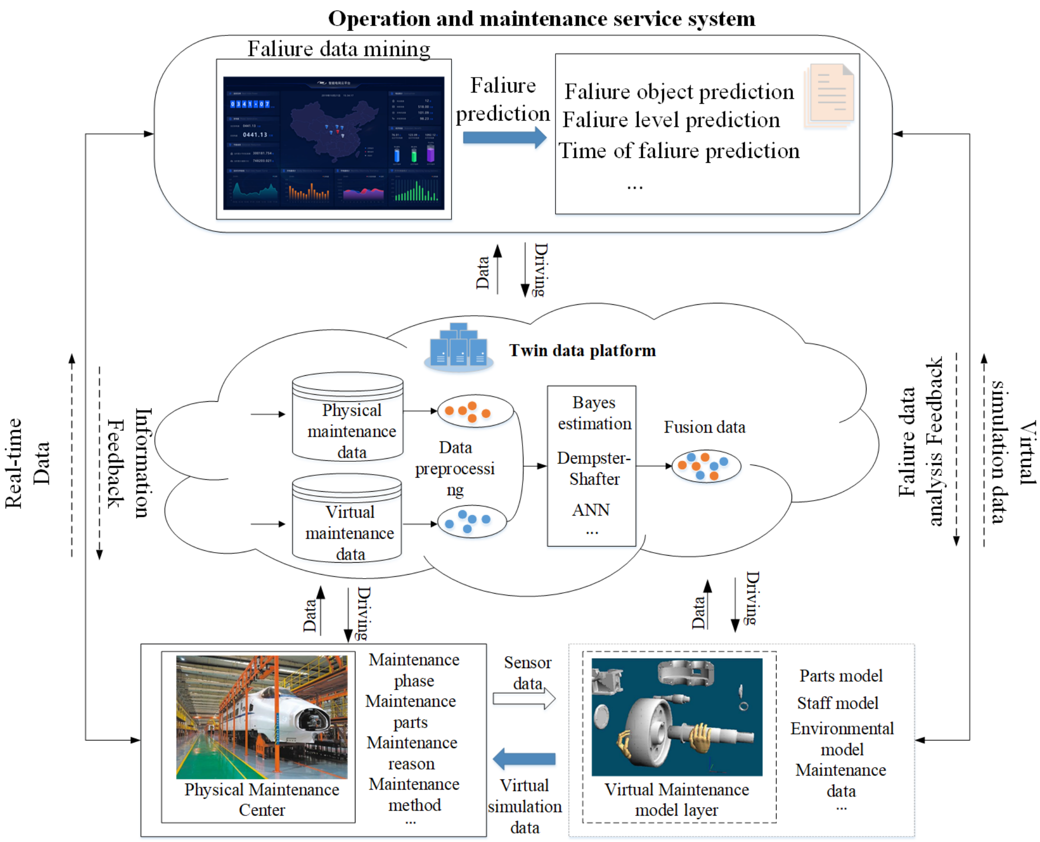

Fig. 1. Intelligent O&M model of complex products based on digital twin

2.1 O&M Service System

OMS is an integrated service platform that encapsulates data, models and algorithms. The subsurface of the system is converted from the resources, such as data, mathematical models, algorithms, O&M data management systems, and O&M data visualization. The data includes real-time data detected by the product operation status sensor, the O&M information simulated by the virtual O&M center, and different data processing algorithms. The system includes an O&M data management system and an environmental information system. The O&M service system selects the applicable algorithms and mathematical models for computing. It provides specific services for the physical operation and maintenance center as well as the virtual O&M center according to their needs. If there are exceptions in the provided services, the algorithm will be reassembled and calculated until the requirements are met. The system can dig out fault information, and then transmit it to the virtual O&M center for verification. The verified results are fed back to the physical O&M center for implementation, providing an essential basis for decisions in O&M.

2.2 Physical Maintenance Center

The physical O&M center is the place for physical product failure maintenance. It obtains the required real-time data mainly through the sensors or embedded modules deployed on the product. The on-site real-time maintenance information of the product flows from the site, mainly including various O&M resources, O&M plans, O&M processes and O&M environments, such as maintenance personnel, equipment and sites.

2.3 Virtual Maintenance Model Layer

The O&M virtual model layer is a digital model reconstruction and mapping of the physical O&M center.A virtual O&M model is built based on the digital twin development platform, complex product operation and maintenance tools. To ensure the high fidelity of the virtual model in the digital twin system, it is necessary to construct the geometric model of the virtual O&M center, define the physical components, set the behavior, and establish the rules. First, SolidWorks, CATIA, UG, and more are used to build a 3D model. Then, the physical parameters of the 3D model, such as the size, shape and materials of the product are set. Based on this, the simulation is carried out by a finite element model. Finally, different prediction algorithm models are established to realize the prediction and evaluation of product O&M status. Integrating the above processes can build a virtual O&M center model. It provides an objective, dynamic and accurate progress status for product O&M, mainly including the allocation of O&M resources, the repair progress and status of each component, and the steps taken by personnel. The staff can dynamically track and monitor the real-time operational status of the O&M center of the physical entity. The O&M virtual model layer can simulate, optimize, predict and validate the O&M process and state of the physical entity in the O&M layer.

2.4 Twin Data Platform

The twin data platform is a data carrier that drives the O&M digital twin model, storing the data generated by each module in the model. In this platform, the data with different scales, dimensions, and granularity need to be compared, correlated and combined to eliminate information silos, so as to provide more comprehensive, consistent, accurate and comprehensive information for each module. The data mainly includes the data generated by the physical O&M center and the virtual O&M model layer, the interaction data between the virtual layer and the physical entity layer, the O&M service system data and the O&M knowledge base. The data in the physical O&M center includes the real-time operation data collected by sensors in the actual operation of products, the failure data, and the maintenance data during the maintenance process. The data of the O&M virtual model layer includes simulation data, model optimization data, fault prediction data and verification data. The interaction data between the physical entity layer and the virtual model layer refers to the data generated in the interaction between the physical entity O&M center and the virtual O&M model layer. The O&M knowledge base is provided by the O&M experts and contains the facts and data related to rules.

3 O&M Mode of Complex Products

3.1 O&M Process of Complex Products

During the operation of a complex product, the sensors collect the operational real-time data input to the twin data platform. The twin data platform processes data on demand. The processed O&M data and professional O&M knowledge are uploaded from the knowledge base into the O&M service system. In the O&M service system, mathematical models and algorithms are used to reason about fault prediction. Combined with the O&M knowledge base, the corresponding preventive maintenance plan is formulated, and the preventive maintenance plan is transferred to the virtual O&M layer for simulation. The simulation results are verified and transferred to the physical O&M shop. According to the information provided by the O&M service system, the O&M resources are rationally allocated, and the complex products are maintained according to the plan. The real-time information of the physical O&M workshop is updated to the twin data platform, and the O&M knowledge base in the data platform is updated to provide adequate data support for the subsequent O&M decisions. At the same time, relevant fault prediction mechanisms are established based on the updated O&M information in the O&M service system, and fault association rules are mined to improve preventive maintenance efficiency. A diagram of the O&M model for a complex product driven by the digital twin is shown in Fig. 2.

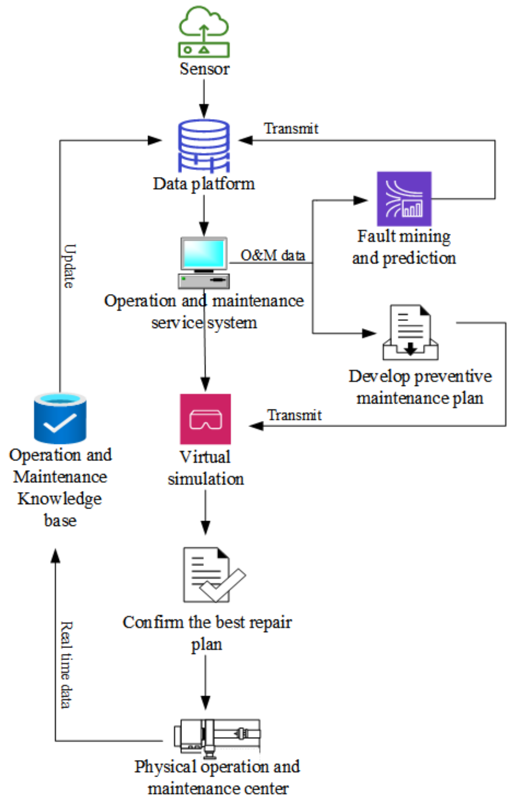

Fig. 2. Complex product O&M operation mode driven by digital twin

3.2 Implementation Route of O&M of Complex Products

The purpose is to reduce the incidence of abnormal events during the product O&M process, so as to improve the security of product operation process. Failure data should be analyzed promptly to identify potential factors affecting failures and provide better decisions for subsequent maintenance. The route of intelligent O&M of complex products based on digital twin is shown in Fig. 3. When an exception or warning occur during the operation of a complex product, the abnormal real-time data collected by the sensor and the fused data in the data platform are uploaded into the O&M service system. The fused data refers to O&M simulation data, product life cycle data, material inventory data and expert O&M knowledge base. It includes data transmission, synthesis, filtering, correlation and synthesis from different information sources to help environmental determination, resource planning, plan verification, and fault diagnosis in intelligent O&M. The fused data will transfer, synthesize, filter, correlate and synthesize the information given by different information sources, so as to obtain valuable data to assist environmental determination, resource planning, plan verification and fault diagnosis in intelligent O&M. In this way, the O&M service system can analyze and mine the historical data to develop the corresponding preventive maintenance plan, and allocate the required O&M resources according to the O&M plans. Preventive maintenance plans and O&M resource allocation are simulated in the virtual simulation layer for scenarios. If the simulation results do not meet the O&M requirements, the simulation data and the O&M plan data will be checked and corrected to obtain an accurate O&M plan. The improved O&M plan is transferred into the physical entity O&M layer to guide the preparation of O&M. Validated preventive maintenance plans are transferred into the virtual model layer, which predefines the O&M process and then gives the manipulation instructions to the physical O&M shop. The physical and virtual O&M workshops processes should be synchronized and the data consistency between the virtual and physical layers should always be maintained. If the results are inconsistent, determine whether the main reason of the inconsistency is the physical entity layer or the virtual model layer. If it is the virtual model layer, the conditions in the virtual model are modified by fusing the data. If it is the physical entity layer, the site is changed or adjusted in time by pulling the data from the platform. After the complex product completes O&M, the historical O&M data is obtained from the actual O&M process, such as the optimal O&M plan and resource allocation, which are stored in the knowledge base. New O&M knowledge can be obtained through data mining when the product runs abnormally the next time. The O&M process simulation can be saved to play back the historical O&M situation. The knowledge base can be updated, and the whole process is a continuous iterative optimization process. Fig. 3 shows a diagram of the operational mechanism of complex product O&M driven by the digital twin.

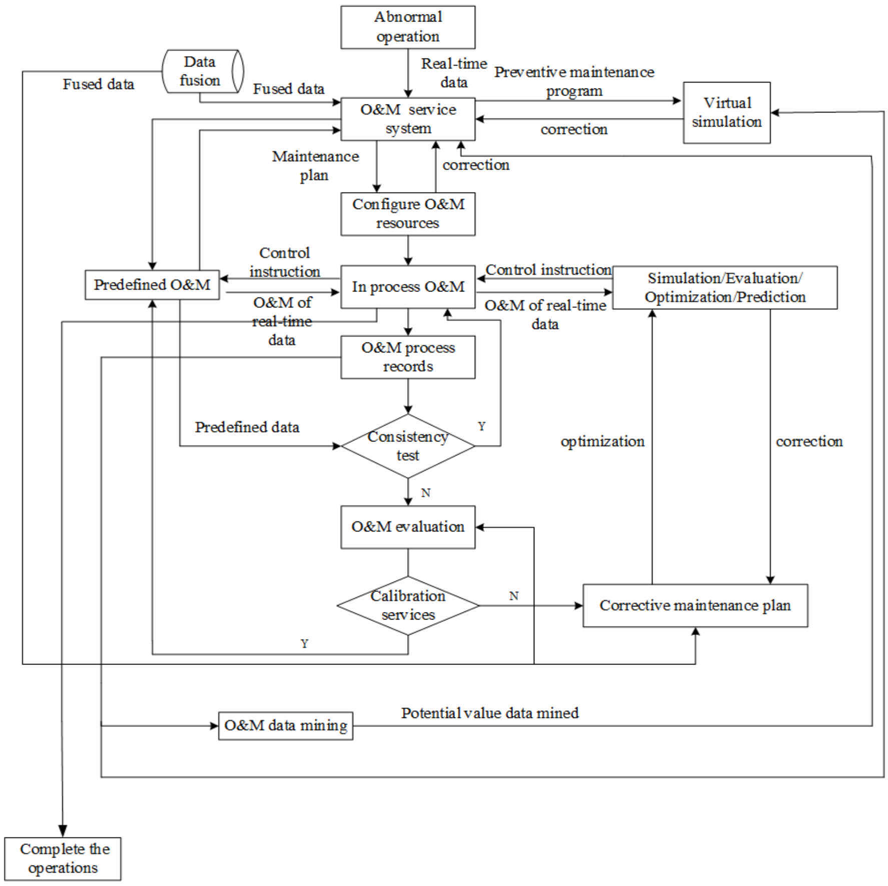

Fig. 3. Digital twin-driven complex product operation and maintenance operation mechanism

3.3 Interoperability of Intelligent O&M Process For Complex Products

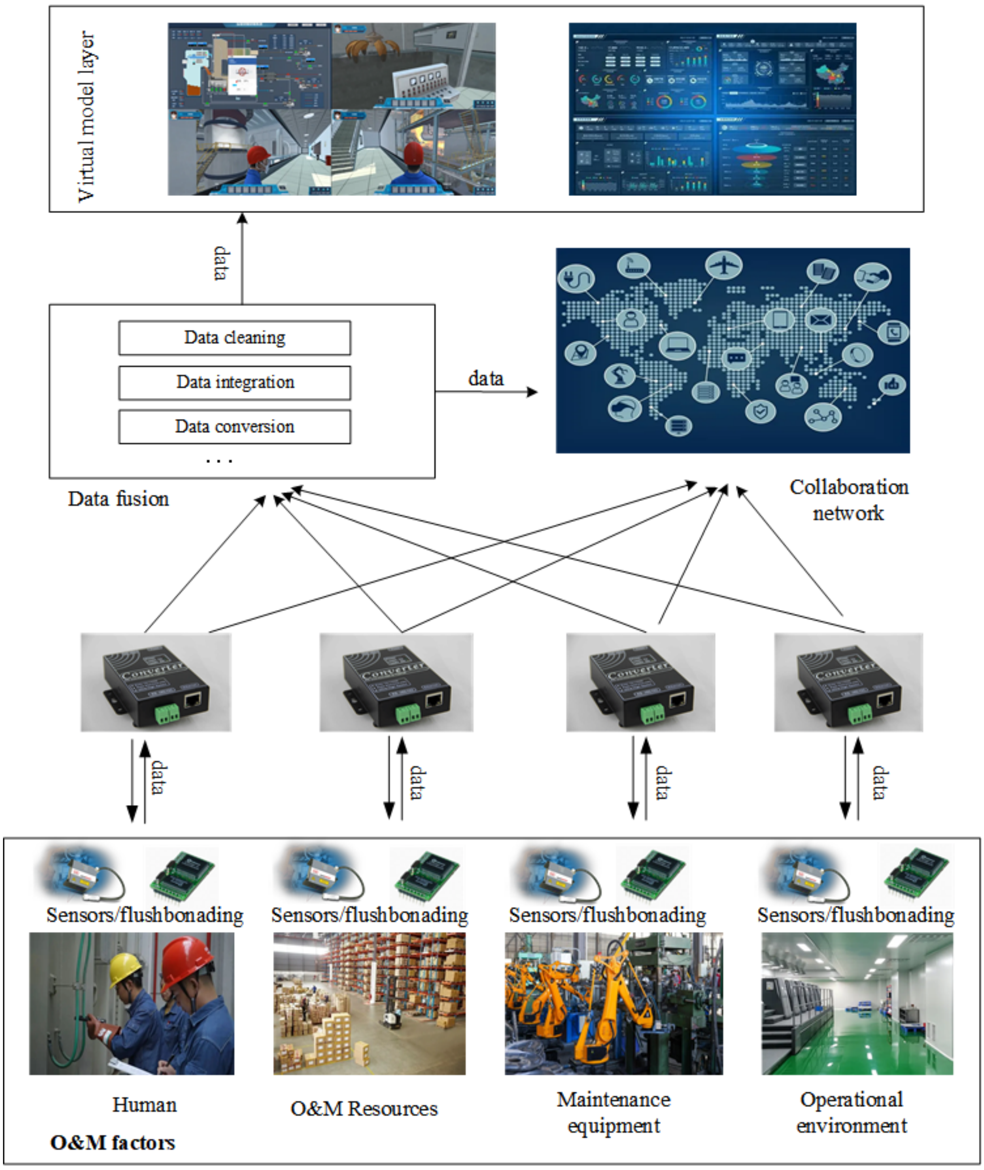

Fig. 4. Intelligent O&M interoperability of complex products based on the digital twin

In the entire O&M process, a large number of sensors are deployed on complex products and O&M equipment to collect information with multi-source heterogeneity. This will cause the virtual O&M layer to be unable to uniformly access data, so that the interoperability between the virtual layer and the physical layer cannot be guaranteed. It is thus necessary to deploy a unified access module to realize data integration in the O&M process. First, the multi-source heterogeneous data is converted into a unified access interface and communication protocol to realize a unified access of data. Then, the semantic ambiguity and structural heterogeneity of these multi-source heterogeneous data are eliminated and converted into uniformly formatted data. Finally, data cleaning is used to remove invalid and missing values to correct errors to ensure data consistency. Based on the comprehensive data, a collaborative network is established to describe different elements and their relationships in the O&M process. Each O&M element can sense, calculate and interact. The integrated data transmits the demand information to a unified access module through the O&M collaboration network. According to the conversion of data access interface and communication protocol, the real-time, accurate and comprehensive data driving for intelligent O&M

process of complex products is provided. Based on digital twins, comprehensive data awareness of intelligent O&M is realized to ensure the integrity and availability of data information. The digital twin technology fully maps the O&M process of complex products and the surrounding environment information to the virtual world, thereby realizing real-time perception of the O&M process of complex products. In the virtual world, the O&M plan is continuously optimized and adjusted to obtain optimal O&M results, which are transmitted to the real O&M center through a comprehensive perception network for product maintenance. Intelligent O&M interoperability of complex products based on the digital twin is shown in Fig. 4.

4 Case Validation

Taking a certain type of EMU bogie system as an example, the intelligent O&M model of complex products based on digital twin is applied to the O&M of EMU bogies, so as to solve the current problem of passive and inefficient O&M of EMU bogies. Combined with the failure correlation analysis results of key components of EMU bogies and the failure prediction times of key components, the information is directly fed back to the O&M department. According to the results, the O&M department sends the maintenance plan to the maintenance centre, and the accurate maintenance plans and scheduling of maintenance resources are prepared in a timely manner.

4.1 Fault Correlation Analysis of EMU Bogies

Failure analysis is the basis of maintenance quality guarantee. The failure analysis results can also provide a basis for the scheduling of maintenance resources to improve the maintenance efficiency. In the O&M process of EMU, to formulate accurate maintenance plans in a timely manner, potential information causing the failure should be excavated to ensure that the factors causing failures can be accurately identified. Here the Apriori algorithm is used to effectively mine the fault association rules and improve the fault diagnosis ability of EMU.

Data Preprocessing. The repair and maintenance records of a CRH series EMU in one year is collected, in which the data of service work and routine security inspection need to be removed. The Pandas function reads the collected fault repair data based on the Python platform. It used the Pandas function to remove the data group where the missing data is located-deleting some useless information, such as notification number, vehicle group number and fault feedback personnel, and filtering out the valuable information for fault diagnosis. The remaining data are sorted with categories and converted into a form that was easy to analyze. The fault information related to bogie components is filtered out, and 51 fault repair data (recorded as $D_{1,} D_{2,} ..., D_{51.}$) are obtained, as shown in Table 1.

Table 1. Partial repair data after pre-processing

Number	Fault location	Failure date	Fault phenomenon	Fault cause	Fault place	Failure number
1	Transmission system	Autumn	Hot axle	Lack of oil in bearings	HeFei	3
2	Transmission system	Spring	Hot axle	Roller breakage	Tangshan	2
3	Transmission system	Autumn	Gear temperature measurement alert	Friction or packing too tight	Jinan	5
4	Transmission system	Summer	Wheel pair failure	Wear of wheel treads	Xi’an	3
5	Transmission system	Winter	Wheel pair failure	Wheel rim cracks	Xi’an	1
...	...	...	...	...	...	...
49	Transmission system	Spring	Oil leakage	Gearbox oil leakage	Chongqing	2
50	Transmission system	Autumn	Wheel pair failure	Spoke plate hole cracks	Zhengzhou	1
51	Transmission system	Autumn	Oil leakage	Shock absorber oil leakage	Chengdu	4

K-means Clustering Analysis. To deeply analyze the data, the fault data are first clustered. In view of the characteristics of EMU fault maintenance data, 51 groups of faults are clustered in the data pre-processing stage. The merged hierarchical cluster is used for analysis. K-means clustering classifies all faults into 3 to 6 classes, and the clustering results are evaluated by the contour coefficient Si of the clusters number k. To obtain the desired K values, the clustering effect is first evaluated using the contour coefficients.

(1) The dataset is divided into k clusters $\{M_1, M_2, \dots, M_k\}$, and k cluster centroids $(\mu_1, \mu_2, \dots, \mu_k)$ are randomly selected. The category of each sample x_i is calculated to minimize the square error.

$$E = \sum_{i=1}^{k} \sum_{x=M_i} \left\| x_i - \mu_j \right\|_2^2 , \tag{1}$$

where μ_j is the cluster center of M_i clusters. The centroid of j is recalculated, and the calculations are repeated until convergence.

$$\mu_j = \frac{1}{\left| M_j \right|} \sum_{x=M_j} x_i . \tag{2}$$

(2) The silhouette coefficient of the i-th element is calculated as follows:

$$S_i = \frac{m_i - n_i}{\max(m_i, n_i)} , \tag{3}$$

where n_i is the average distance between the i-th element and other elements in the same cluster. A cluster m outside the cluster is selected to calculate the average distance between all elements in m and the element. The minimum value m_i, $m_i = min\{m_{i1}, m_{i2}, \dots, m_{ij}\}$ is found.

When $n_i < m_i$, $S_i = \frac{1-n_i}{m_i}$; when $n_i = m_i$, $S_i = 0$; when $n_i > m_i$, $S_i = \frac{n_i}{m_i - 1}$. The closer S_i is to 1, the better the clustering effect.

The calculated silhouette coefficients corresponding to different cluster numbers k are shown in Table 2.

Table 2. The correspondence between the cluster number and silhouette coefficient

The cluster number k	Silhouette coefficient S_i
3	0.358
4	0.472
5	0.597
6	0.461

As shown in Table 2, the effect is optimal when the fault data is clustered into 5 categories. Combined with the actual fault conditions of EMU bogie components, the cluster analysis results of 51 sets of fault data are shown in Table 3.

Table 3. Fault clustering results (k = 5)

Serial number	Failure sample	Failure phenomenon
Fault 1	$D_3, D_6, D_7, D_{10}, D_{11}, D_{12}, D_{15}, D_{24}, D_{26}, D_{37}, D_{45}$	Gear temperature measurement alert
Fault 2	$D_1, D_2, D_9, D_{16}, D_{17}, D_{20}, D_{23}, D_{25}, D_{29}, D_{32}, D_{33}, D_{40}, D_{47}$	Axle failure

Fault 3	$D_{18}, D_{27}, D_{34}, D_{35}, D_{43}, D_{48}$	Component collision
Fault 4	$D_{13}, D_{19}, D_{22}, D_{36}, D_{41}, D_{44}, D_{49}, D_{51}$	Equipment oil spill
Fault 5	$D_4, D_5, D_8, D_{14}, D_{19}, D_{21}, D_{28}, D_{30}, D_{31}, D_{38}, D_{39}, D_{42}, D_{46}, D_{50}$	Wheelset failure

Analysis of Apriori Fault Association Rules. To better analyze the correlation between the faults, and provide a data basis for subsequent O&M decisions, the Apriori algorithm is used to mine association rules for fault data.

Apriori algorithm is a data mining algorithm that mines the frequent item set of association rules. It mainly searches for implicit association relationships between attributes from large-scale data sets to solve the association analysis problem of large-scale data sets. The core idea is to analyze the data stored in the data platform and find the implied patterns between different data sets and their strong correlations. Among them, the form of relationship is divided into two types of frequent item set association rules. Frequent item sets are attribute sets that frequently appear together. Association rules imply that there may be a strong relationship between two features. Both frequent item sets and association rules needs to calculate the support and confidence. The support indicates the probability that the attribute value *X* under a particular dimension appears in the dataset D. The confidence indicates the likelihood that two different attributes *X* and *Y* appear in the same dataset. In addition, there is an essential concept of lift, which indicates the availability of a metric association rule. When *lift* > 1, the higher indicates a higher degree of correlation. When *lift* < 1, the lower indicates a lower degree of correlation. When *lift* = 1, it indicates that there is no correlation. The following equation can be expressed:

$$support(\mathrm{X}) = \frac{count(\mathrm{X})}{count(D)} = P(X) . \tag{4}$$

$$confidence(X \to Y) = \frac{\sup port(X \cup Y)}{\sup port(Y)} = P(X \mid Y) . \tag{5}$$

$$lift(X \to Y) = \frac{confidence(X \to Y)}{support(X)} . \tag{6}$$

In order to obtain ideal association rules, it is important to determine the support and confidence. The number of rules generated under different support and confidence levels is different. According to the analysis, the threshold of support is set to 0.6%, and that of confidence is 40%. The main steps of the algorithm are shown in Table 4.

Table 4. Algorithm process

Algorithm process:	
Input:	Bogie component failure maintenance database *D*; Minimum support; Confidence threshold;
Output:	Association rule set *R* in database *D*.
1.	C_1=find_frequent_1-items set(D);//Scan D to find all item sets C_1 with frequent 1
2.	$L_1 = \{c \in C_1 \mid c.count \geq minsupport\}$;// Remove the unmet items from C_1 according to the minimum support level to obtain frequent item set L_1
3.	for{k=2;L_{k-1}≠Null; k++};
4.	$C_1 = sc\_\ candidate\ (L_{k-1})$;
5.	if the value of an attribute in C_i is not in['fault cause']and not in ['fault phenomenon'];
6.	delete the frequent item;
7.	for each t in *D*;
8.	$C_1 = subset\ (C_k, t)$;
9.	for each candidate $C \in C_i$;
10.	$c.count = c.count + 1$;

```
Lk = {c ∈ Ck | c.count ≥ minsupport};
for each item in Lk;
if item attribute value not in [‘ fault cause ‘];
delete the frequent item;
Return L;
R=generate_big_rules(L);
Return R;
```

There are 76 strong correlation rules. The former items of the rule contain the fault phenomenon and the latter items of the rule are the cause of the fault, as shown in Table 5.

Table 5. Fault association rules for bogie components

Rule number	Association rule results			Support	Confidence	Lift
1	The opposite side of the gearbox wheel seal failed, Summer	→	Gear temperature measurement warning	7.8%	100%	12.75
2	Poor lubrication condition, Bearing failure	→	Hot axle shaft	5.8%	60%	10.2
3	Gearbox leakage	→	Gearbox failure/damage	7.8%	100%	12.75
4	Abnormal profile of rail head	→	Shake the car	5.8%	100%	17
5	Gearbox failure/damage	→	Gear temperature measurement warning	7.8%	100%	12.75
6	Too many people	→	Wheel diameter super poor	5.8%	100%	17
...	...	→	...	...	...	....
71	Abnormal noise, Bearing failure	→	Sensor fault	5.8%	100%	17
72	Gear box collision	→	Degree sensor fault	5.8%	100%	17
73	Oil is mixed with impurities	→	Hot axle shaft	5.8%	100%	17
74	Bearing failure	→	Gear temperature measurement warning	5.8%	100%	17
75	Wheel pair collision, Heating of bearing	→	Bent axle	5.8%	100%	17
76	Xi’an, Foreign objects pressed into the wheel tread	→	Wheel tread crack	5.8%	60%	10.2

Based on the cluster analysis, the fault maintenance data of key components of EMU bogies are mined and analyzed, and the potential laws among multiple factors are identified. The obtained results are as follows:

(1) According to rules 1, 3, 5, and 74, association rules with high support can be obtained, indicating that the high-frequency fault gearbox temperature warning is often caused by the failure of the wheel-to-side seal of the gearbox and the damage of the gearbox in summer. Wheel tread damage is often accompanied by abnormal wheel noise and brake shoe locking. Thus the temperature of the gearbox should be paid more attention in summer. When the wheel makes abnormal noise, it is also necessary to timely check whether the brake shoes are locked. More attention should be paid to the damage of the tread, replacing the wheel in time to prevent the occurrence of failure.

(2) According to rules 2 and 73, the main cause of axle heating is the change of lubrication conditions. It is necessary to check whether the lubricating oil meets the lubrication conditions.

(3) According to the rules 4 and 6, the reason for the shaking caused by the vehicle may be that the diameter of the wheel is out of tolerance and the profile of the rail head is abnormal. Wheels and rails require regular inspection and maintenance.

(4) According to rules 71 and 72, when the sensor fails, it is necessary to check whether the bearing is damaged and whether the gearbox is subject to collision. When a general sensor failure requires the parts being replaced, these parts may need to be replaced at the same time.

(5) According to rule 75, the bending of the axle may be caused by an external force on the wheel set. At the same time, the lubrication of the bearings also needs to be checked.

4.2 Failure Prediction of Key Components of Bogie

Fault prediction is the premise of analyzing fault causes. According to the fault prediction results, combined with the fault association rules, the fault location and cause can be efficiently determined, and the corresponding solutions can be formulated. In addition, it can also better schedule O&M resources to ensure the smooth development of EMU O&M. The gray prediction method is used to predict failure and the failure number of key components of the bogie in the next year. The advantage of this method is that it can generate irregular raw data to obtain a generated sequence with strong regularity. The specific modeling process and algorithm steps are as follows:

Step 1: Create the original dataset;

$$x^{(0)} = \left\{x^{(0)}(1), x^{(0)}(2), x^{(0)}(3), x^{(0)}(4), \ldots, x^{(0)}(n)\right\}. \tag{7}$$

Step 2: Generate a new sequence by accumulating the original data sequence;

$$x^{(1)} = \left\{x^{(1)}(1), x^{(1)}(2), x^{(1)}(3), x^{(1)}(4), \ldots, x^{(1)}(n)\right\}. \tag{8}$$

$$x^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(i), k = 1, 2, \ldots, n \,. \tag{9}$$

Step 3: Build the data matrix;

$$B = \begin{bmatrix} -\frac{1}{2}\left[x^{(1)}(1) + x^{(1)}(2)\right] & 1 \\ -\frac{1}{2}\left[x^{(1)}(2) + x^{(1)}(3)\right] & 1 \\ -\frac{1}{2}\left[x^{(1)}(3) + x^{(1)}(4)\right] & 1 \\ \ldots & \ldots \\ -\frac{1}{2}\left[x^{(1)}(n-1) + x^{(1)}(n)\right] & 1 \end{bmatrix}, \; Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ x^{(0)}(4) \\ \ldots \\ x^{(0)}(n) \end{bmatrix}. \tag{10}$$

Step 4: Obtain parameters a and u by the least square method;

$$\hat{a} = \begin{bmatrix} a \\ u \end{bmatrix} = (B^T B)^{-1} B^T Y \,. \tag{11}$$

Step 5: Discrete the corresponding equation;

$$\hat{x}^{(1)}(k+1) = (x^{(0)}(1) - \frac{u}{a})e^{-ak} + \frac{u}{a} \,. \tag{12}$$

Step 6: Obtain the predicted values.

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k) \,\cdot \tag{13}$$

The occurrence times of five kinds of failures of EMU bogies in recent 20 years are counted, as shown in Table 6. After the calculation of the gray prediction model, the failure prediction results for the next year are shown in the Table 7:

Table 6. Number of failures in 20 years

Time/Year	Gear temperature measurement alert	Axle failure	Component collision	Equipment oil spill	Wheelset failure
2001	20	36	22	19	57
2002	32	34	20	15	52
2003	34	30	19	10	47
2004	32	27	16	12	39
...	...	...	...	...	...
2020	15	12	8	10	22

Table 7. Failure prediction results

No.	The fault phenomenon	Predictive value
1	Gear temperature measurement alert	14
2	Axle failure	10
3	Component collision	7
4	Equipment oil spill	8
5	Wheelset failure	19

According to the results, it can be seen that the frequency of wheel set faults is the highest. According to the results of fault correlation analysis, it is necessary to pay attention to the cause of the faults. To ensure the normal operation of EMU in the next stage, it is necessary to formulate a maintenance plan in advance.

4.3 Operation Mechanism of Intelligent O&M of EMU Bogie Wheelset Based on Digital Twin

The operation mechanism of intelligent O&M of EMU bogie wheel set based on digital twin drive is shown in Fig. 5. In the O&M service system, the correlation analysis and prediction of fault data are realized by intelligent algorithms. According to the prediction results, combined with the cause of failures, the maintenance plan for the bogie wheel set is formulated in advance, and the spare parts required for the wheel set maintenance are dispatched in time to ensure the smooth progress of the maintenance.

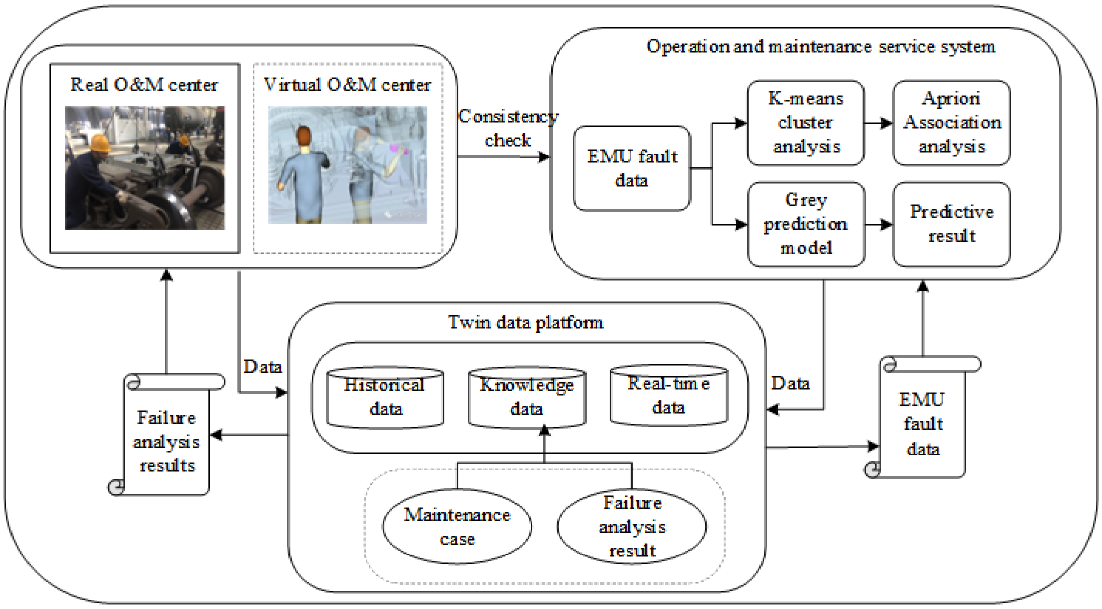

Fig. 5. Intelligent O&M mechanism of EMU bogie wheel set based on digital twin

(1) The O&M service system analyzes and predicts the fault data of EMU wheelset operation, and feeds the data back to the twin data platform. The twin data platform provides fault prediction results, fault analysis results, and maintenance case base for the O&M virtual simulation center. According to the data analysis, it can be seen that the factors causing the failure of the wheel set may include brake shoe locking, coupling oil leakage, wheel diameter out of tolerance and more. After receiving these data, the virtual O&M center checks and determines the cause of failure in the virtual space in advance. Then combined with actual data, it formulates a preliminary O&M plan to simulate the O&M process. In the virtual O&M center, the plan will be continuously revised according to real-time data to determine optimal maintenance plan.

(2) After an optimal maintenance plan is determined, the information is sent to the real O&M center to check whether the O&M resource conditions meet the implementation of the plan. If the conditions are met, the maintenance will be simulated in the virtual space. If not, the quantity information of the required bearings and couplings will be sent to the supplier, so as to arrange the replenishment of spare parts in time.

(3) The optimal maintenance plan, resource scheduling plan, and real O&M data of the physical O&M center generated in the O&M of the wheelset are stored in the knowledge base, so that the next time the same fault occurs, the failure can be handled efficiently. According to the fault correlation analysis and fault prediction results, a pre-maintenance solution for the O&M process is provided. Through the virtual-real interaction of the digital twin, the intelligent O&M of dynamic update of O&M knowledge, real-time diagnosis, and reasonable resource scheduling are realized. The intelligent O&M framework of complex products based on digital twin proposed in this paper is verified.

5 Conclusion

Taking the intelligent development requirements of the manufacturing industry as the research background, this paper aims to realize the intelligence and collaboration of the O&M process. The application of digital twin technology in the complex product operation and maintenance process is studied, and the realization of digital twin interoperability is discussed. The main contributions of this paper are reflected in the following aspects:

(1) Based on the analysis of the shortcomings of the existing O&M management model, a complex product intelligent O&M framework based on digital twin technology is proposed.

(2) In order to realize the intelligent management and control of the O&M process, a digital twin intelligent O&M mechanism based on the status of O&M resources is designed. In the O&M process, it can provide design methods and implementation ideas for the integration of virtual and real interactions, dynamic updating of O&M knowledge and the active execution of O&M services. The interoperability achieved by the digital twin is illustrated.

(3) Taking the EMU bogie as an example, the K-means algorithm, Apriori algorithm and the gray prediction model are used to comprehensively analyze the failure data of EMU bogie. Combined with the digital twin-based intelligent O&M operation mechanism proposed in this paper, the data analysis results illustrate the feasibility and effectiveness of the model.

According to introducing the digital twin technology into the O&M process of complex products, the intelligent management of O&M is realized to provide a new perspective for the future intelligent O&M development of complex products. As a theoretical research, this paper only verifies the application of the framework model and data-driven O&M knowledge acquisition based on the theoretical framework and technical implementation. Recently, the application of digital twin is still in the stage of theoretical exploration, and there are still some key technologies that need to be broken through. In the future, we will focus on the specific modeling and data fusion technology of the digital twin of all elements of the product.

Acknowledgements

This work was funded by the General Project 51974229 of the National Natural Science Foundation of China, and the Shaanxi Province Innovation Capability Support Program2021TD-27.

References

[1] Z.J. Wu, Z.B. Gao, Y.L. Cao, X.P. Ye, J.X. Yang, Tolerance Design and Adjustment of Complex Customized Product Based on Cloud Manufacturing, Procedia CIRP 27(2015) 169-175.

[2] Y.R. Wang, Y. Li, W.Z. Ren, Digital twin-based research on the prediction method for the complex product assembly abnormal events, International Journal of Computer Integrated Manufacturing 34(12)(2021) 1382-1393.

[3] S. Ren, L.C. Shi, Y. Liu, W.H. Cai, Y.F. Zhang, A personalised operation and maintenance approach for complex products based on equipment portrait of product-service system, Robotics and Computer-Integrated Manufacturing, 80(2023) 102485.

[4] G.B. Gao, D.M. Zhou, H. Tang, X. Hu, An Intelligent Health diagnosis and Maintenance Decision-making approach in Smart Manufacturing, Reliability Engineering & System Safety 216(2021) 107965.

[5] C.C. Liu, H.H. Zhu, D.B. Tang, Q.W. Nie, T. Zhou, L.P. Wang, Y.J. Song, Probing an intelligent predictive maintenance approach with deep learning and augmented reality for machine tools in IoT-enabled manufacturing, Robotics and Computer-Integrated Manufacturing 77(2022) 102357.

[6] P. Yepez, B. Alsayyed, R. Ahmad, Intelligent assisted maintenance plan generation for corrective maintenance, Manufacturing Letters 21(2019) 7-11.

[7] K. Antosz, L. Pasko, A. Gola, The Use of Intelligent Systems to Support the Decision-Making Process in Lean Maintenance Management, IFAC-PapersOnLine 52(10)(2019) 148-153.

[8] T.R. Mohan, J.P. Roselyn, R.A. Uthra, D. Devaraj, K. Umachandran, Intelligent machine learning based total productive maintenance approach for achieving zero downtime in industrial machinery, Computers & Industrial Engineering 157(2021) 107267.

[9] T.-E. Li, Z. He, Kernel-based fuzzy clustering algorithm for multi-faults diagnosis of complex products, System Engineering Theory and Practice 33(1)(2013) 181-186.

[10] E. Ruschel, E.A.P. Santos, E.D.F.R. Loures, Mining Shop-Floor Data for Preventive Maintenance Management: Integrating Probabilistic and Predictive Models, Procedia Manufacturing 11(2017) 1127-1134.

[11] T. Yang, X.L. Yi, S.W. Lu, K.H. Johansson, T.Y. Chai, Intelligent Manufacturing for the Process Industry Driven by Industrial Artificial Intelligence, Engineering 7(9)(2021) 1224-1230.

[12] A.K. Sharma, R. Bhandari, C. Pinca-Bretotean, C. Sharma, S.K. Dhakad, A. Mathur, A study of trends and industrial prospects of Industry 4.0, Materials Today: Proceedings 47(10)(2021) 2364-2369.

[13] G. Zhang, C.-H. Chen, X.G. Cao, R.Y. Zhong, X.Y. Duan, P.F. Li, Industrial Internet of Things-enabled monitoring and maintenance mechanism for fully mechanized mining equipment, Advanced Engineering Informatics 54(2022) 101782.

[14] The State Council, The notice of made in china 2025 strategy of the state council. <http://www.gov.cn/zhengce/content/2015-05/19/content_ 9784.htm>, 2015 (accessed 25.11.2019).

[15] F. Psarommatis, G. May, A standardized approach for measuring the performance and flexibility of digital twins, International Journal of Production Research 61(20)(2023) 6923-6938.

[16] R. Rosen, G.V. Wichert, G. Lo, K.D. Bettenhausen, About the importance of autonomy and digital twins for the future of manufacturing, IFAC-PapersOnLine 48(3)(2015) 567-572.

[17] M. Betty, Digital Twins, Environment and Planning B: Urban Analytics and City Science, SAGE Journals 45(5)(2018) 817-820.

[18] F. Tao, J.F. Cheng, Q.L. Qi, M. Zhang, H. Zhang, F.Y. Sui, Digital twin-driven product design, manufacturing and service with big data, The International Journal of Advanced Manufacturing Technology 94(9-12)(2018) 3563-3576.

[19] F. Tao, M. Zhang, J.F. Cheng, Q.L. Qi, Digital twin workshop: a new paradigm for future workshop, Computer Integrated Manufacturing Systems 23(1)(2017) 1-9.

[20] C. Zhuang, J. Gong, J. Liu, Digital twin-based assembly data management and process traceability for complex products, Journal of Manufacturing Systems 58(Part B)(2021) 118-131.

[21] O. Davies, A. Makkattil, C. Jiang, M. Farsi, A Digital Twin Design for Maintenance Optimization, Procedia CIRP 109(2022) 395-400.

[22] S. Li, J. Wang, J. Rong, W. Wei, A digital twin framework for product to-be-designed analysis based on operation data, Procedia CIRP 109(2022) 179-184.

[23] Z. Wang, W. Jia, K. Wang, Y. Wang, Q. Hua, Digital twins supported equipment maintenance model in intelligent water conservancy, Computers and Electrical Engineering 101(2022) 108033.

[24] A. Coraddu, L. Oneto, F. Baldi, F. Cipollini, M. Atlar, S. Savio, Data-driven ship digital twin for estimating the speed loss caused by the marine fouling, Ocean Engineering 186(2019) 106063.