ISSN: 2664-9640

Research on a Self-enhancing Maintenance Knowledge Base Method based on Fault Causal Association Probabilities

Linjiao Hou, Bin Liu

National-recognized Enterprise Technology Center, Sunward Equipment Co., Ltd., Changsha, Hunan 410100, China

Abstract

To address the problems of over-reliance on expert experience, knowledge fragmentation, and low diagnostic efficiency in the intelligent maintenance of widely distributed complex equipment, this paper proposes an intelligent maintenance method based on a self-enhancing dynamic knowledge base. First, a structured integration framework for multi-source heterogeneous maintenance knowledge is established to achieve the systematic organization of fault cases and service records. Second, a "high-incidence fault probability model" based on statistical analysis of historical fault data is proposed, forming a priority-based diagnostic strategy of “high-incidence faults first, low-incidence faults later . Furthermore, a large language model is integrated to enable natural language interaction and accurate knowledge retrieval. In addition, a knowledge credibility evolution and closed-loop optimization mechanism based on maintenance log feedback is introduced, driving the knowledge base to evolve from static storage into a dynamic self-enhancing system. On this basis, a progressive knowledge architecture consisting of “Level 1: authoritative knowledge, Level 2: knowledge to be verified, and Level 3: AI-derived knowledge”, together with its self-enhancing interaction workflow is designed. The proposed method is further validated through practical after-sales maintenance scenarios covering three stages: work-order submission, fault analysis, and case accumulation. The proposed method is further validated through practical after-sales maintenance scenarios covering three stages: : work-order submission, fault analysis, and case accumulation . The results show that the proposed method can significantly improves the overall efficiency of fault description, localization, and troubleshooting, thereby providing a systematic solution with continuous self-enhancing capability for the intelligent maintenance of complex electro-mechanical-hydraulic integrated equipment.

Keywords

Intelligent Maintenance; Fault Diagnosis; Knowledge Base; High-incidence Fault Probability Model; Large Language Model; Closed-loop Learning.

1. Introduction

In recent years, after-sales service models have undergone profound transformation, driven primarily by the shift from "passive maintenance" to "active value-added" services, as reflected in service contract–oriented operations and lifecycle customer engagement. . This transformation has led to comprehensive upgrades in both technical capabilities and management paradigms .Service response mechanisms have evolved from the traditional reactive workflow of “customer repair request–work order dispatch–on-site maintenance” toward proactive warning and intelligent scheduling systems [1][2]. Technical methods have likewise evolved from reliance on “experienced technicians and paper-based manuals” to remote intelligent diagnostics integrating Internet of Things (IoT) monitoring, big-data driven prediction, and augmented reality (AR)-assisted maintenance technologies [3][4][5][6]. Spare

ISSN: 2664-9640

parts management has advanced from inefficient states characterized by simultaneous inventory backlog and shortages toward intelligent coordination and high-turnover systems driven by demand forecasting [7][8].

However, although most enterprises have established relatively complete service workflows and accumulated a large number of fault cases and service records, their service operations still heavily depend on expert experience. Problems such as excessive manual data entry, prolonged spare-parts procurement cycles, and lengthy service response periods remain prevalent . Overall, these enterprises are still at the early stage of digitalization and continue to exhibit a significant gap with respect to the industry's development objectives for intelligent service and maintenance systems [5].

To address the aforementioned challenges, this paper aims to obtain large volume of fault cases and service records from internal enterprise sources, including cloud-based maintenance service platforms, as well as from extensive external information sources. By analyzing the characteristics of these multi-source heterogeneous data, potential association rules among fault phenomena, fault locations, and root causes are identified,, thereby automatically expanding the case data volume of the maintenance knowledge base. The ultimate goal is to develop an intelligent maintenance knowledge base system with data self-enhancement capability, providing an innovative solution for the intelligent operation and maintenance of construction machinery.

2. Methodology

2.1. Technical Approach

The core technical approach of this study is the construction of a fault causal association probability model. The core criterion of this model is that "the factor with the highest probability is regarded as the most likely fault cause or fault location," thereby establishing a quantitative high-frequency fault model. On one hand, this model can directly guide on-site troubleshooting, by prioritizing the inspection of high-probability components or fault locations.. On the other hand, through continuous accumulation of diagnostic results, the model can reversely validate and expand case data in the knowledge base, thereby enabling data-driven enhancement of maintenance knowledge [5].

Traditional fault analysis and diagnosis method primarily is based on fault-tree-based hierarchical troubleshooting. The fault tree method, which sequentially checks various possible fault causes and fault locations based on the physical structure and spatial configuration of the faulty equipment, is typically time-consuming and labor-intensive.

Experience-based analysis derived from historical data can be employed to identify the most probable fault causes and fault locations, thereby eliminating the need for exhaustive step-by-step inspection and significantly improving fault-handling efficiency. The core concept of this analysis method is that, rather than investigating all potential fault causes with equal priority, historical data are utilized to calculate the association probabilities between different fault phenomena and specific fault locations or root causes. . The system then automatically recommends the fault cause and location with the highest probability as the primary inspection targets , forming an efficient diagnostic process of "prioritizing high-incidence fault locations before low- incidence ones." The author terms this analysis method as the "highest fault probability fault cause/location identification method."The probability-based fault cause/location identification method is based on big data technology, achieving precise identification, scientific decision-making, and dynamic evaluation through data collection, analysis, and modeling.

ISSN: 2664-9640

2.2. Data Sources

Through the self-enhancing dynamic knowledge base in the big data platform, historical maintenance records are collected, including information such as "fault phenomenon," "fault locations," " corrective solutions," "equipment models," "operating hours," and other required information. These data are subsequently structured and standardized for further analysis.

2.2.1. Data Processing and Analysis Methods

Statistical analysis: Statistical analysis techniques are used to mine massive volumes of maintenance data and to calculate the occurrence frequencies of different fault locations under specific fault phenomena. Model construction: Association probabilities are calculated based on the statistical frequencies , forming a diagnostic probability map in the form of "Phenomenon A → Location X (75% probability), Location Y (20% probability), Others (5% probability)." This map is the knowledge core of the "fault cause and location identification " model.

A “high-probability diagnostic checklist” is established to determine high-frequency fault probabilities, thereby constructing a fault cause and fault location identification model.

Subsequently, the model is utilized to support intelligent fault diagnosis. During the fault analysis process, the system prioritizes the invocation of the “high-probability diagnostic checklist” generated by the model, thereby guiding engineers or customers to perform targeted troubleshooting. This mechanism can substantially reduce the average fault diagnosis and troubleshooting time.

2.2.2. Establishing "High-Probability Diagnostic Checklist,"

A “high-probability diagnostic checklist” is established to determine high-frequency fault probabilities, thereby constructing a fault cause and fault location identification model. First, the above "fault cause and location identification " model is used to upgrade the knowledge base, forming a self-enhancing dynamic knowledge base. This model is the core rule for promoting Level 2 knowledge to Level 1 knowledge. If a fault association (e.g., “phenomenon–location”) in the Level-2 knowledge base has been repeatedly retrieved and validated as “effective” through user feedback, and its accumulated credibility reaches a predefined threshold, the model identifies the association as a confirmed fault cause and fault location. Consequently, the corresponding knowledge is automatically upgraded to high-confidence Level-1 knowledge. Then, this model is used to guide intelligent diagnosis. During fault analysis, the system calls the "high-probability diagnostic checklist" generated by this model to guide engineers or customers in targeted troubleshooting, thereby greatly reducing the average fault diagnosis and troubleshooting time.

2.3. Construction of Self-Enhancing Knowledge Base Structure

With the continuous growth of business operations, the knowledge base in the enterprise cloud-based maintenance service platform has become increasingly enriched. . To achieve self-enhancing management, , this paper organizes the knowledge base into a three-level progressive structure as shown in Figure 1.

ISSN: 2664-9640

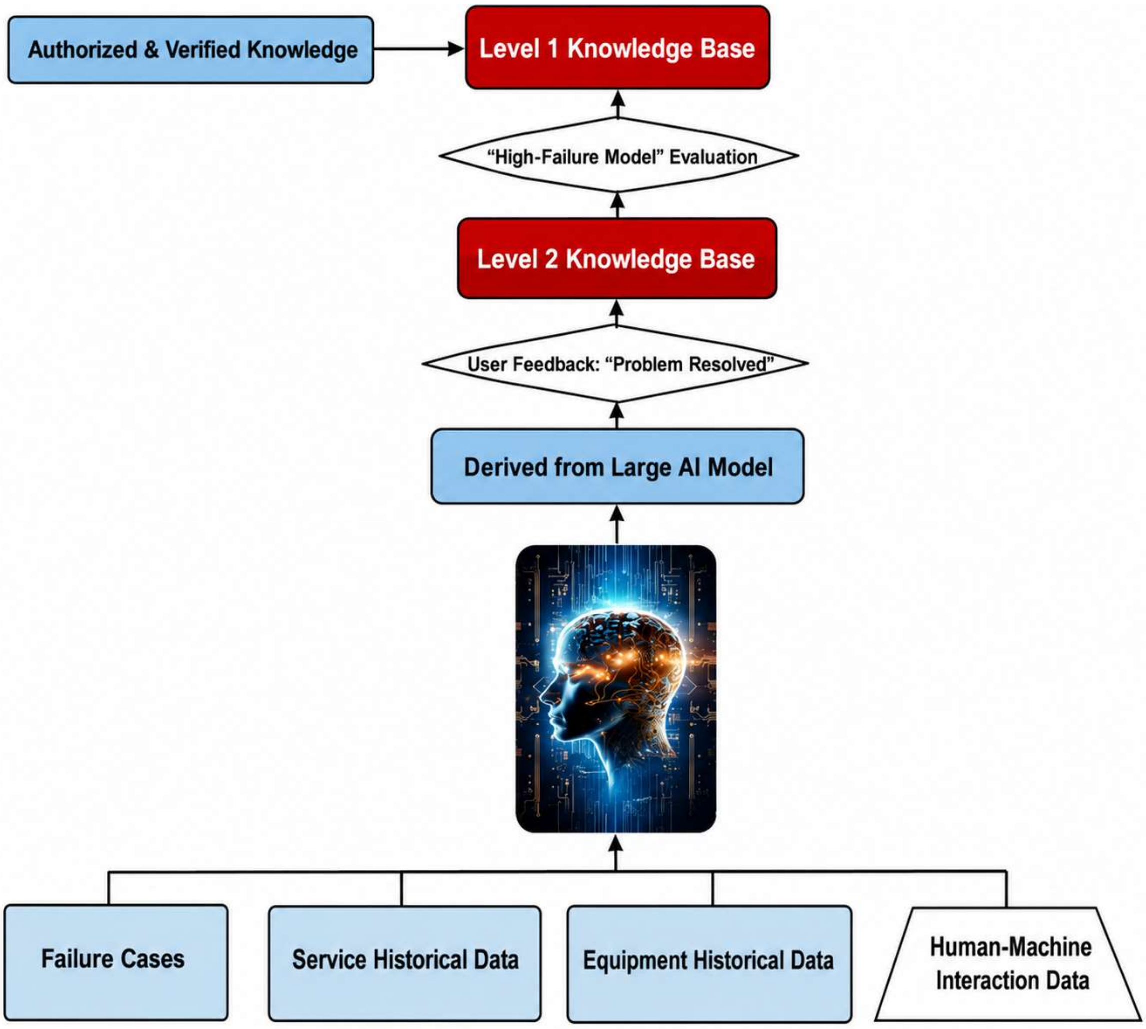

Figure 1. Knowledge Base Accumulation Process

Level 1 Knowledge Base (Authoritative Knowledge): This level consists of high-confidence solutions that have been reviewed by domain experts or confirmed through multiple on-site validations with high credibility. Such knowledge is prioritized during intelligent fault diagnosis and troubleshooting processes. Level 2 Knowledge Base (Knowledge to be Verified): This level contains knowledge knowledge derived from historical work orders, maintenance records, and related service data. Although such knowledge possesses certain reference value, it has not yet undergone rigorous validation.. Such knowledge is available for reference but must be marked "to be verified."Level 3 Knowledge Base (AI-derived Knowledge): This level consists of candidate solutions generated by large language models through logical reasoning and innovative expansion based on Level 1 knowledge, Level 2 knowledge, and general-domain knowledge. Such knowledge serves as the primary source for knowledge-base data enhancement and must undergo rigorous feedback-based validation before being promoted to higher knowledge levels [3].

2.4. Human-Computer Interaction Self-Enhancing Knowledge Base Logical Reasoning Process

This system aims to achieve efficient human-computer interaction and self-evolution of the knowledge base through hierarchical knowledge retrieval combined with AI intelligent generation. The core workflow follows the principle of "retrieval first, follow by generation, with feedback-driven enhancement."

ISSN: 2664-9640

The workflow proceeds as follows::

Step 0 – Workflow Initiation:

Step 0, Process Trigger: The human-computer interaction system receives the user's query and activates the AI intelligent engine [3].After the workflow is triggered, the system first enters the retrieval process of the Level-1 Knowledge Base (Authoritative Knowledge):

Step 1 – Retrieval:

The system first searches the Level 1 Knowledge Base (containing high-credibility and verified knowledge).

Step 2 – Success:

If a related case is retrieved, the AI reasoning engine is called to perform precise matching and logical inference , directly returning a reliable answer to the human-computer interaction system.

Step 3 – ValueThis path provides authoritative, credible solutions, offering the best user experience.

Then enter the Level 2 Knowledge Base (knowledge to be verified) retrieval process:

Step 1, Retrieval: If no relevant result is found in the Level-1 Knowledge Base , retrieve the Level 2 Knowledge Base (containing initially accumulated, knowledge need to be verified or "high-incidence fault model" cases).

Step 2, Success: If relevant information is retrieved, call the AI reasoning engine for in-depth analysis and reasoning.

Step 3, Return: The generated result is returned with an AI-generated content disclaimer, indicating that the information is for reference only and has not yet been fully validated.

Then enter the AI large model derived knowledge (innovative expanded knowledge) generation process:

Step 1, Activation: If no relevant result is found in either the Level-1 or Level-2 Knowledge Base , activate the AI large model's self-generation capability.

Step 2, Generation: Generate a new solution based on general knowledge, problem context, and the "high-incidence fault model" theory.

Step 3, Return: Return the generated results with the same prompt: "Based on AI-generated content, for reference only."

Finally enter the knowledge base self-enhancing closed-loop (credibility evolution) generation process:

Step 1, Feedback Collection: The human-computer interaction system collects user feedback on the final outcome of this interaction.

Step 2, Positive Enhancement: If the user confirms "problem solved," the system automatically increments the "credibility" indicator of the knowledge entry retrieved (whether from Level 2 knowledge base or AI model generation) by +1.

Step 3, Closed-loop Value: Through continuous feedback, high-value knowledge gradually promotes from "Level 2 Knowledge Base" to "Level 1 Knowledge Base," while ineffective knowledge is marginalized, achieving dynamic optimization and self-enhancement of the knowledge base.

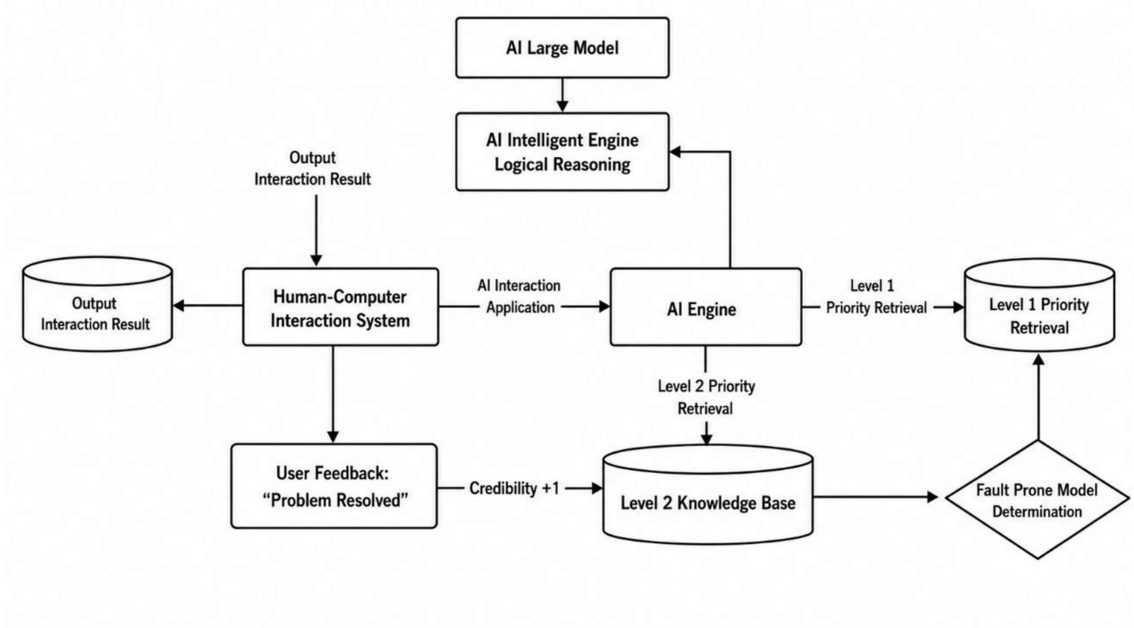

Figure 2. Knowledge Base Self-Enhancing Interaction Process

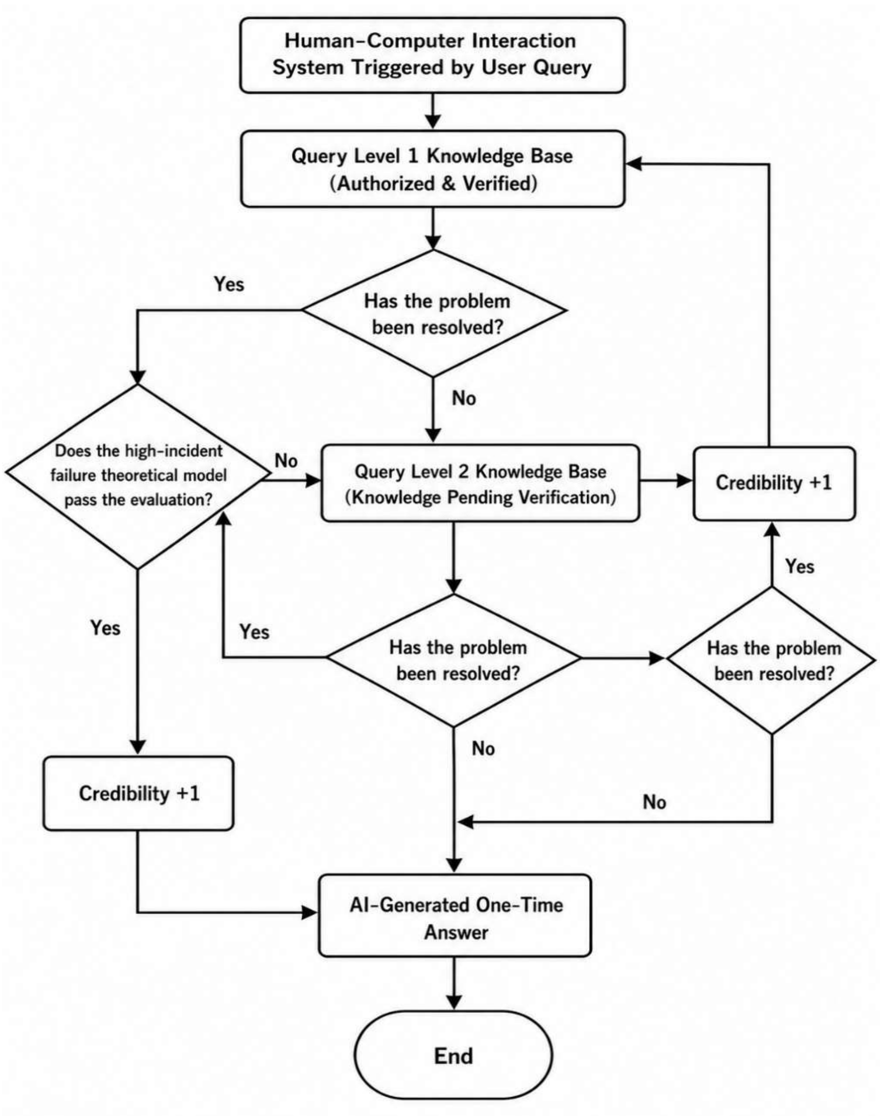

Figure 3. Self-Enhancing Knowledge Base Logical Reasoning

2.5. Data Source Analysis of the Self-Enhancing Knowledge Base

As mentioned above, the data source of the self-enhancing knowledge base is maintenance service data. The maintenance service process includes the following three stages:

Customer Initiates Work Order Submission: The customer initiates a service work order through the human-computer interaction system interface. The system guides the customer to further improve the fault description based on the maintenance knowledge base and attempts self-service troubleshooting until the problem is solved or the work order is submitted. After submission, the system automatically generates an "Intelligent Maintenance Work Order" and an "Intelligent Maintenance Plan."

Fault Analysis Process: The service engineer receives the "Intelligent Maintenance Work Order" and the automatically generated "Intelligent Maintenance Plan" submitted by the customer and begins the fault analysis process. The engineer performs fault troubleshooting according to the "Intelligent Maintenance Plan," and the system provides further inspection prompts based on the maintenance knowledge base until the problem is solved.

Fault Case Knowledge: The AI large model analyzes human-computer interaction logs, historical fault records, fault cases, etc., as derived knowledge. When the derived knowledge solves a problem, it enters the Level 2 knowledge base. When the credibility reaches the requirement of the "high-incidence fault model," the data enters the Level 1 knowledge base as a data source for the self-enhancing knowledge base.

Figure 4. After-sales Maintenance Process Diagram

The following figure shows the knowledge base process during after-sales maintenance.

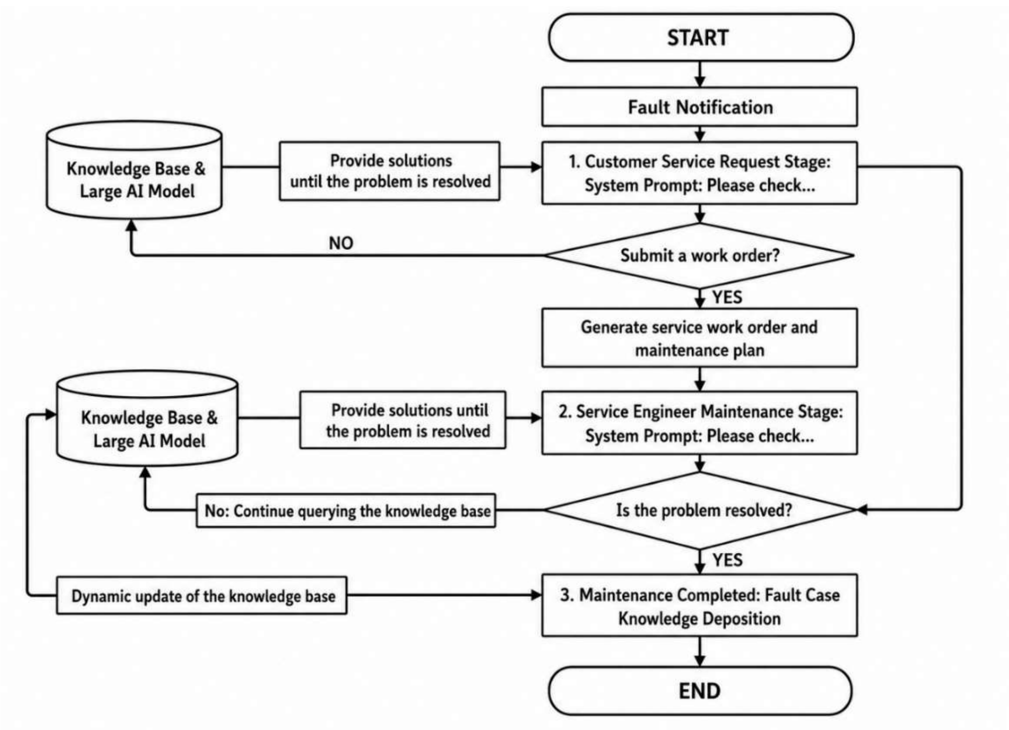

Figure 5. Knowledge Base Call Flowchart during After-sales Maintenance

ISSN: 2664-9640

2.6. Construction of Fault Causal Association Probability Model to Obtain High-Incidence Fault Model Criteria

2.6.1. Definitions

Definition 1, Equivalent Record: Let C be an attribute on relation instance r, and t and t' be records in r. If t and t' have the same value on attribute C, i.e., t[C] = t'[C], then records t and t' are said to be equivalent with respect to attribute C. For example, equipment operation logs of devices with the same equipment type, brand, engine, and controller can be considered equivalent records.

Definition 2, Equivalence Set: Let C be an attribute on relation instance r, and t be a record in r. The set of all records in r equivalent to t with respect to attribute C is called the equivalence set with respect to t, denoted as [t]C, [t]C = {t' | t' ∈ r, t[C] = t'[C]}. The number of records in [t]C is denoted as |[t]C|. For example, the set of equivalent records of equipment operation logs for devices with the same equipment type, brand, engine, and controller.

Definition 3, Credibility: For association rule C → FC, within each equivalence set under attribute C, ti, tj ∈ r(U) (i, j = 1, 2, … N) are arbitrary records, N is the number of records in the database. The total number of records not satisfying Condition 1 within each equivalence set is denoted as Err, Err = |C → FC|. The credibility of this fault space can be denoted as Con(C), calculated as follows: Con(C) = (1 - Err/N) × 100%. Credibility is used to measure the probability of fault reasoning. For example, credibility of a lock-up fault: with the same equipment type, brand, engine, controller, if the bus is abnormal, lock-up occurs, probability 99%.

Definition 4, Fault Reasoning: Let r be an instance of relation table r(U), X#, FC ∈ U, C ⊆ U - {X#, FC}. If records t in the equivalence set with respect to C in r satisfy: if t[X#] = False, then t[FC] = False; when Con(C) > MinC, fault reasoning holds, denoted as FD({X#} ∪ C) → FC.

Definition 5, Strong Correlation Fault Reasoning: FD({X#} ∪ C) → FC, when Con(C) = 1, it is a strong correlation fault model. This paper aims to mine strong correlation fault reasoning. The setting of the minimum credibility MinC is domain-related, requiring interaction with users and setting by domain experts. If MinC is set too small, the mined fault correlations will be weak correlations. If MinC is set too large, originally strong correlation fault models are easily missed.

2.6.2. Algorithm

The objective of this algorithm is to identify all meaningful high-frequency fault probability models. The main idea of the algorithm is to employ a hierarchical breadth-first search strategy on a given relation instance r, exploring all attributes and their combinations associated with r. Ultimately, the algorithm outputs interesting attribute sets that satisfy the predefined confidence requirements.

Input: Relation instance r, all attribute set U of r, minimum credibility MinC.

Output: C → FC. If C ∈ ABF, then there exists an interesting fault model C → FC, output in the form ("C", FC, "Con").

Step 1: ABF = ∅, Levelset = ∅; D = U – FC

Step 2: Level = 1; Levelset = {{a} | a ∈ D}

Step 3: while (Levelset ≠ ∅)

Step 4: for each element C in Levelset do

Step 5: if (Con(C) > MinC) then ABF = ABF ∪ C

Step 6: if (Con(C) = 1) then Levelset = Levelset – C

Step 7: end for

Step 8: NextLevelSet = ∅

Step 9: for all s, k in Levelset do

ISSN: 2664-9640

Step 10: if |s ∪ k| = Level + 1 then Nextlevelset = NextLevelset ∪ {s ∪ k}
Step 11: end for
Step 12: Levelset = NextLevelset; Level = Level + 1
Step 13: end while
Step 14: Return ABF

The output of this algorithm satisfies the attribute C of interesting fault models. Credibility represents the reliability of the fault model, and whether attribute C is interesting is guaranteed by this indicator. Algorithm Step 5 and Step 7 are pruning rules that reduce the number of searches. Appropriately setting the credibility threshold can yield more interesting fault associations.

3. Results Case Study

To illustrate the implementation process of the fault cause/location discrimination model and evaluate its implementation effect, this paper takes the excavator "engine bog-down" fault as an example, comparing the traditional method with the "high-incidence fault probability" model algorithm for fault cause/location troubleshooting.

(1) Traditional Method

According to the traditional "standardized fault diagnosis logic," troubleshooting for "engine bog-down" requires organizing fault troubleshooting knowledge. Such knowledge will continuously enter the knowledge base, and AI software will be developed to automatically organize and maintenance knowledge base documents.

The traditional troubleshooting process is:

Step 1: Customer initiates a service work order, describing the issue as:: “The excavator operating speed has decreased, and black smoke is being emitted.”

Step 2: Service engineer receives the work order and describes the fault: On-site, the engine sounds muffled, accompanied by significant black smoke. Or It is Slow, weak, and stuck. Check the monitor; there are fault codes or warning icons such as engine overload, hydraulic overheating, intake pressure. Hydraulic oil temperature rises rapidly. Engine automatically stalls; restarting is normal, but stalls again when operate is performed. The excavator body may even tremble.

Step 3: Fault localization: The service engineer determines on-site that the fault location may be engine bog-down.

Step 4: Fault cause finding. The steps for finding engine bog-down causes are as follows:

Step 1, Check and replace the air filter.

Step 2, Check and replace the diesel filter, remove water and impurities in the fuel line.

Step 3, Read fault codes to see if the ECU has records.

Step 4, Observe exhaust smoke color (if black smoke is severe, it usually indicates insufficient intake or fuel injection problems).

Step 5, Check hydraulic oil quality and temperature.

Step 6, Listen to engine sound changes under different loads.

Step 7, Measure engine speed at no-load and load to see if the drop is too large.

Step 8, Use a pressure gauge to measure main pump output pressure and pilot pressure.

Step 9, Check turbocharger intake pressure.

Step 10, Perform injector and cylinder pressure tests.

Step 11, Check the main pump regulator and proportional solenoid valve.

Step 12, Check the resistance and signal voltage of all related sensors.

Step 13, Finally consider engine overhaul (cylinder pressure test) or main pump disassembly overhaul.

(2) Application of the "High-Incidence Fault Probability" Model Algorithm

Following the above traditional method, 13 steps are required to identify the fault cause. Such fault troubleshooting is redundant time-consuming and labor-intensive.

Using the "fault cause discrimination model," historical data in the self-enhancing dynamic knowledge base shows:

80% of cases are caused by "engine cylinder damage";

15% of cases are caused by "filter clogging";

5% are other causes.

Then, the "fault location discrimination model" generates the following diagnostic priority: First check the "engine cylinder" (high-incidence location); Second check the "filter"; Finally troubleshoot other possibilities.

Accordingly, the model automatically generates a diagnostic priority checklist: first recommend focusing on checking the "engine cylinder"; if that cause is ruled out, then recommend checking the "filter"; finally troubleshoot other possibilities. If the fault happens to be the high-incidence "engine cylinder damage," the engineer only needs 1-2 targeted checks to locate the problem, eliminating 11 ineffective troubleshooting steps compared to the traditional method, greatly improving diagnostic efficiency. Even when the fault originates from a low-probability “other” category, the system still provides a clear probability-ranked troubleshooting pathway, avoiding unfocused and exhaustive inspection procedures.

4. Conclusion

The essence of the "fault cause/location" model based on the "high-incidence fault probability" algorithm is a data-driven dynamic probability model, which enables the self-enhancement capability of the dynamic knowledge base through continuous learning and evolution: This represents a transformation from experience-based knowledge to data-driven knowledge; that is, converting the vague experiential judgment of senior technicians-such as “this problem is most likely caused by failure at that component”-into quantifiable and computable probabilistic data.

1) Transformation from experience-based form to data-driven knowledge . That is, converting the vague experiential judgment of senior technicians-such as “this problem is most likely caused by failure at that component”-into quantifiable and computable probabilistic data.

2) This also represents a transformation from static data to dynamic data. With the continuous feedback of new maintenance data, the model probabilities are updated in real time, and the "high-incidence" map evolves accordingly. As a result, the knowledge base and diagnostic system acquire self-learning and continuous self-enhancement capabilities.

3) The fault cause and location identification model is not only a tool for improving single diagnostic efficiency but also the core engine driving the entire intelligent maintenance system to achieve closed-loop evolution and continuous optimization.

4) Through continuous feedback from maintenance logs and the knowledge credibility evolution system, the proposed method effectively promotes the transformation of the knowledge base from a static storage repository into a dynamic self-enhancing system. . The application case shows that this method can significantly improve the overall efficiency of fault description, localization, and troubleshooting.. Future research will further explore model generalization capabilities across equipment models, as well as finer-grained credibility decay and update mechanisms, providing theoretical support for building a more robust and intelligent complex equipment maintenance system.

ISSN: 2664-9640

References

[1] Dai, A. S., Yang, D., Luo, Z., & He, S. G. (2025). The digital and intelligent operation and maintenance mechanism of high-end equipment driven by information technology – A multi-case study. Industrial Engineering and Management, 30(5), 206–218.

[2] Liu, B. M. (2025). Intelligent scheduling of remote communication and control for electromechanical power distribution. Smart China, 28(12), 44–45.

[3] Min, Y., Wang, Y. Q., Ma, X., & Dong, Y. N. (n.d.). Intelligent fault diagnosis in hybrid chemical processes under limited samples based on multi-feature fusion learning. Computers & Chemical Engineering, 207(1), 960–968.

[4] Guan, H. Y. (2025). Research on construction and optimization of intelligent diagnosis model for remote oil pipeline leakage. Computer Knowledge and Technology, 21(33), 106–108.

[5] Zhang, X. J. (2025). Remote operation and maintenance and fault diagnosis methods for smart manufacturing equipment driven by digital twins. Electronic Technology, 54(11), 13–15.

[6] Sun, S. J. (2025). Analysis of fault diagnosis and intelligent maintenance technology for electromechanical equipment. Electronic Technology, 54(11), 166–167.

[7] Kumar, L., & Khedlekar, U. K. (2025). A hybrid AI and blockchain framework for intelligent supply chain optimization with reinforcement learning-based inventory management, stochastic demand forecasting, and sustainable supplier selection. Unidentified Journal, 30(12), 1–29.

[8] Li, Z. Q., & Zuo, T. T. (2025). Integrated model of intelligent predictive inventory management considering data fluctuations. Computer Simulation, 42(6), 220–224, 378.