122 КОМП'ЮТЕРНІ НАУКИ ТА ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

UDC 004:37:001:62 DOI: 10.31498/2225-6733.50.2025.336233

MODEL FOR DETERMINING SIMILARITY BETWEEN CASES IN FAILURE DIAGNOSIS OF SHIP POWER PLANT EQUIPMENT

Vychuzhanin V. *D.Sc. (Engineering), professor, Odesa Polytechnic National University, Odessa, ORCID: https://orcid.org/0000-0002-6302-1832, e-mail: v.v.vychuzhanin@op.edu.ua;*

Vychuzhanin A. *PhD (Computer science), associate professor, Odesa Polytechnic National University, Odessa, ORCID: https://orcid.org/0000-0001-8779-2503, e-mail: vychuzhanin.o.v@op.edu.ua*

Modern ship power plants (SPPs) are high-tech systems that involve the continuous collection and analysis of large volumes of diagnostic data. Improving the efficiency of technical operation of such systems requires the development of intelligent diagnostic tools capable of accurate recognition and interpretation of typical failures. This paper addresses the problem of assessing similarity between cases within the framework of Case-Based Reasoning (CBR), which is a relevant approach under conditions of limited training data and high variability of failures. A comprehensive review of current similarity assessment methods is provided, including metric-based approaches, aggregation operators (e.g., Choquet integral), logic-based inference systems, and semantic graph models. The main limitations of these methods are identified, such as low adaptability to heterogeneous data, poor interpretability, and strong dependence on a priori knowledge. The authors propose a unified similarity assessment model that accounts for the heterogeneity of diagnostic features. The model integrates different similarity metrics – Euclidean, cosine, and Jaccard – depending on the type of data (numerical, categorical, or multi-valued). The diagnostic significance of features is represented by weight coefficients, which are optimized using numerical methods. The paper also presents a formal description of the case structure, a classification of typical failures, and an algorithm for matching based on the operational context. The proposed approach ensures interpretability, adaptability, and scalability of the model, confirming its applicability in intelligent decision support systems for failure diagnosis of SPP equipment. Practical implementation of the model can significantly enhance the reliability, safety, and efficiency of marine systems operation.

Keywords: *intelligent decision support systems, ship power plant, equipment failure, case, Case-Based Reasoning, case similarity, metric, interpretability.*

Statement of the problem

In modern complex technical systems (CTS), timely and accurate fault diagnosis of equipment plays a key role in ensuring uninterrupted operation, safety, and high performance. The increasing complexity and heterogeneity of such systems have led to the widespread adoption of intelligent fault diagnosis methods based on knowledge reuse, among which Case-Based Reasoning (CBR) stands out in particular. Diagnosis using CBR involves comparing the current situation with previously encountered cases, with the effectiveness of this approach largely depending on the accuracy of similarity assessment between cases.

Analysis of the latest achievements on the identified problem

Similarity assessment between cases is a key task in CBR systems, as it directly affects the relevance of the retrieved solutions. Modern approaches range from traditional metrics to machine learning and logic-based models, but they face challenges related to scalability, interpretability, and adaptability to heterogeneous data [1]. For instance, Neykov and Stefanova [2] highlighted the flexibility of rule-based heuristics but noted difficulties in scaling and tuning weights. Chen et al. [3] applied graph neural networks to detect structural similarities, but pointed out their low transparency and complex preprocessing. The use of the Choquet integral for feature aggregation was proposed in [4], though it complicates the formalization of expert preferences. In [5], Ye introduced neutrosophic logic with cotangent-based similarity measures to handle uncertainty, yet interpretability remained an issue. Time series analysis methods such as Dynamic Time Warping demonstrated robustness to distortions [6], but suffer from high computational complexity and limited semantic sensitivity. Adaptive fuzzy systems (El Bitar et al. [7]) have shown resilience to noise but remain limited in industrial validation across diverse datasets.

Traditional approaches insufficiently account for probabilistic fault dependencies, operational conditions, and degradation dynamics. Therefore, there is a need to develop an adaptive and unified similarity assessment methodology capable of effectively comparing fault cases – particularly in the context of ship power plants (SPPs) – by incorporating probabilities, semantics, and operational conditions.

p-ISSN: 2225-6733; e-ISSN: 2519-271X

Purpose and task statement

The aim of this study is to develop a unified and adaptive methodology for assessing the degree of similarity between equipment failure cases within SPPs. This methodology should integrate metric and probabilistic approaches, handle heterogeneous data, and ensure result interpretability for subsequent implementation in intelligent CBR-based diagnostic systems.

The methodology should be based on a similarity metric capable of effectively matching new failures with archived cases and identifying the most relevant analogs. It must take into account both quantitative and qualitative characteristics of failures, support the integration of metric, logical, and learning-based components, and operate with heterogeneous data – including symbolic, numerical, and time-series formats. Interpretability of results for experts and adaptability of the methodology to changing operating conditions of SPPs are also essential.

Summary of the main material

To develop a formalized model for determining the similarity between cases, it is first necessary to structure information about typical failures of SPP equipment [8, 9] and the corresponding diagnostic features. Table 1 presents a classification of failures based on key parameters – temperature, vibration level, pressure, and typical manifestations. This structure makes it possible to identify diagnostically significant features, which are further formalized as case parameters.

To establish the relationship between features and failure types, three scenarios can be identified. Failure A (bearing overheating) is associated with prolonged load, accompanied by an increase in temperature and vibration. Failure B (hydraulic shock in the cooling system) occurs during abrupt mode changes due to unstable pressure and temperature. Failure C (injector defect) is characterized by reduced fuel pressure and unstable engine operation. These examples demonstrate that comprehensive parameter analysis allows for accurate differentiation of failure types and identification of their causes. In the tasks of diagnosing failures in CTS, including SPP, the model of case representation and comparison is a key element of intelligent decision support. Traditionally, similarity between cases is evaluated using Euclidean and Manhattan distances in feature spaces based on the numerical representation of failure parameters [10, 11]. Some studies also consider categorical features, but often without adapting the metrics to their nature [12]. More advanced approaches suggest the use of ontologies or logical representations [13]; however, these require complex verification and are not always robust to incomplete data. The model proposed in this study differs from existing approaches in several key aspects. First, it employs specialized metrics tailored to the type of diagnostic parameter: Euclidean distance for numerical features, cosine similarity for categorical features, and Jaccard distance for multiple-value features, ensuring adequate similarity evaluation. Second, the model incorporates a weighting system that reflects the diagnostic significance of each feature, thereby improving the accuracy and interpretability of the results. Third, it allows for weight optimization using numerical methods, which enables adaptation of the model to real expert assessments and empirical data. Finally, the model is designed to operate with limited training samples, which is particularly important under SPP operational conditions, where the number of recorded failure cases is limited, and the data may be incomplete or heterogeneous.

The overall similarity measure between two cases A and B is defined as a weighted sum of individual similarity coefficients.

For two cases $A_i = (x_1, \ldots, x_n)$ and $B_i = (y_1, \ldots, y_n)$, the similarity $S_i(A,B)$ is calculated using the formula [14]:

$$S(A,B) = \sum_{i=1}^{n} \omega_i \cdot sim_i(A_i, B_i), \qquad (1)$$

where $sim_i(A_i, B_i) \in |0,1|$ – is the partial similarity measure for the i - th feature, and ω_i is the weight of the i-th parameter.

Table 1

Classification of typical failures and diagnostic features

Failure Type	Temperature	Vibration	Pressure	Typical Symptoms
Failure A	>90°C	2.0–3.5 mm/s	>10 bar	Often associated with cooling system overheating
Failure B	75–85°C	1.5–2.5 mm/s	6–9 bar	Related to gradual bearing wear
Failure C	>85°C	2.5–4.0 mm/s	9–11 bar	Observed with unstable fuel supply

The partial similarity functions are defined as follows:

Absolute Difference Similarity (for numerical features normalized to [0,1]):

$$sim_{i.}(A_i, B_i) = \frac{|A_i - B_i|}{\max(A_i) - \min(A_i)}$$

Cosine Similarity (for categorical features, such as failure type, encoded using one-hot encoding) [15]:

$$sim_{i.cos}(A_i, B_i) = \frac{A_i^{\rightarrow} \cdot B_i^{\rightarrow}}{||A_i^{\rightarrow}|| \cdot ||B_i^{\rightarrow}||}$$

Euclidean Similarity (for normalized numerical parameters, such as failure probability or risk category) [16, 17]:

$$sim_{i.euc}(A_i, B_i) = 1 - |A_i - B_i|$$

Jaccard Similarity (for set-based features, e.g., failure subsystems) [18]:

$$sim_{i.euc}(A_i, B_i) = \frac{|A_i \cap B_i|}{|A_i \cup B_i|}$$

For practical implementation in the context of ship engine unit (SEU) diagnostics, the following specification of the similarity measure is used. The formal definition of similarity between cases is implemented using the k-nearest neighbors (k-NN) method, which calculates the similarity between cases. The parameter weights are optimized to reflect their relative importance:

$$S(A, B) = \sum_{i=1}^{n} \omega_i \cdot d_i(A_i, B_i), \quad (2)$$

where ω_i – weight of the i-th parameter;

$d_i(A_i, B_i)$ – dissimilarity measure of the iii-th parameter between cases A_i and B_i;

n – total number of parameters

To improve diagnostic accuracy, the case base is regularly updated and optimized. The case similarity is calculated as follows:

$$S(A_i, B_i) = \alpha_{s.i} \cdot d_{types.i}(A_i, B_i) + \beta_{s.i} \cdot d_{probability.i}(A_i, B_i) + \gamma_{s.i} \cdot d_{components.i}(A_i, B_i), \quad (3)$$

where $d_{types.i}(A_i, B_i)$ – similarity measure of failure types;

$d_{probability.i}(A_i, B_i)$ – difference in failure probabilities;

$d_{components.i}(A_i, B_i)$ – similarity of affected components;

$\alpha_{s.i}$, $\beta_{s.i}$, $\gamma_{s.i}$ – weighting coefficients reflecting the importance of different comparison aspects.

$$d_{types.i}(A_i, B_i) = \begin{cases} 1, if\ the\ failure\ types\ are\ identical \\ 0.5, if\ they\ belong\ to\ the\ same\ category, \\ 0, \text{if they are different} \end{cases}$$

$$d_{probability.i}(A_i, B_i) = 1 - |\mathrm{P}(A_i) - (B_i)|,$$

$$d_{components.i}(A_i, B_i) = \frac{|\mathrm{C}(A_i) \cap \mathrm{C}(B_i)|}{|\mathrm{C}(A_i) \cup \mathrm{C}(B_i)|},$$

where $\mathrm{P}(A_i)$ and (B_i) – are the failure probabilities of components;

$\mathrm{C}(A_i)$ *and* $\mathrm{C}(B_i)$ – are the sets of components involved in the failures. This allows for a quantitative assessment of the overlap between affected subsystems.

For normalization of the overall metric, the following is used:

$$S_{norm}(A_i, B_i) = \frac{S(A_i, B_i)}{\alpha_{s.i} + \beta_{s.i} + \gamma_{s.i}}$$

The coefficients $\alpha_{s.i}$, $\beta_{s.i}$ and $\gamma_{s.i}$ are selected via optimization to minimize classification error. Figure 1 shows a diagram of the influence of $\alpha_{s.i}$, $\beta_{s.i}$ and $\gamma_{s.i}$ on the final similarity score between precedent pairs. The number of the precedent pair is given on the x-axis.

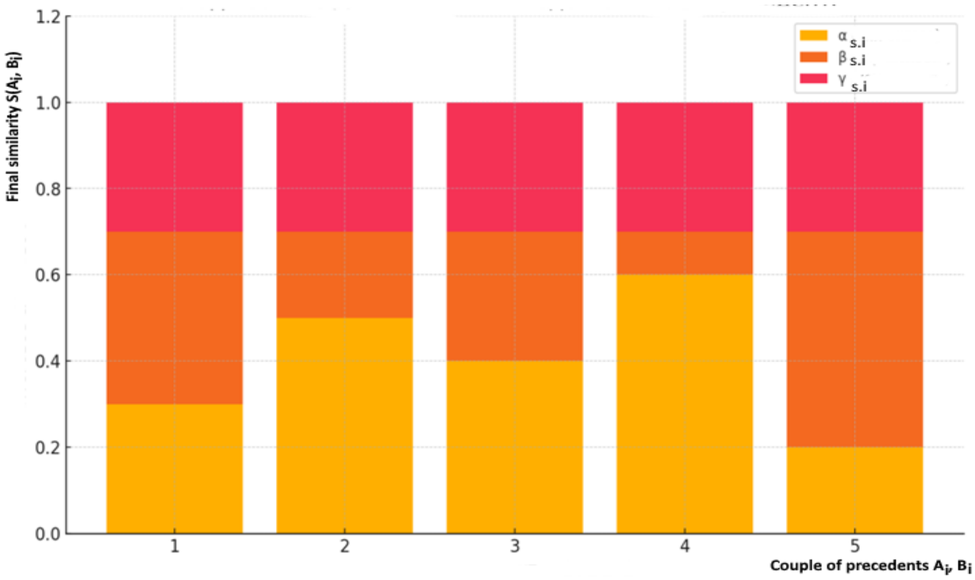

Fig. 1 – Diagram of the influence of $\alpha_{s.i}$, $\beta_{s.i}$ and $\gamma_{s.i}$on the final similarity between precedents

The visualization of the components $\alpha_{s.i}$, $\beta_{s.i}$ and $\gamma_{s.i}$ in the structure of the similarity metric $S(A_i, B_i)$ illustrates their respective contributions to the overall analogy assessment between precedent pairs. The coefficient $\alpha_{s.i}$ reflects the influence of failure type: when the types match, $d_{types.i}(A_i, B_i)$ tends toward 1, amplifying the similarity score even with a moderate weight. In the case of mismatch, the influence drops significantly, which reduces the risk of false positives.The components $\beta_{s.i}$ (related to probability) and $\gamma_{s.i}$ (component overlap) provide smooth compensation in scenarios where failure types differ, enhancing both the interpretability and adaptability of the model. This structure ensures that precedent comparison remains robust to data heterogeneity. Looking ahead, it would be appropriate to integrate probabilistic inference mechanisms to account for prior probabilities and uncertainty.

<u>Approach 1: Probabilistic weighting of parameters.</u>

The weight w_i for a parameter can be defined as a function of the probability of occurrence of the corresponding failure:

$$w_i=f(p_i), \quad (4)$$

where p_i – prior probability of the parameter/failure

<u>Approach 2: Extension of similarity measure through probabilities</u>

$$S(A,B) = \sum_{i=1}^{n} \omega_i \cdot sim_i(A_i, B_i) \cdot P_i(A_i, B_i), \quad (5)$$

where $P_i(A_i, B_i) \in |0,1|$ – the probability of joint occurrence of values A_i and B_i in precedent cases, obtained from empirical data or an expert-defined model.

<u>Approach 3: Bayesian interpretation.</u>

The model may use similarity as evidence in support of a diagnostic hypothesis:

$$P(H_k|observed\ similarity|) \propto P(\text{сходство}|H_k) \cdot P(H_k), \quad (6)$$

where H_k – the hypothesis of membership in a specific failure class.

In CBR-based diagnostic systems, the efficiency and accuracy of precedent matching rely heavily on the correct measurement of their similarity. However, the classical similarity model – based on the weighted sum of partial distances – overlooks a critical factor: the varying importance of feature matches depending on their probabilistic nature and context.

To improve the interpretability and reliability of results, this study employs a modified similarity metric that incorporates a probabilistic multiplier reflecting the frequency of joint feature occurrences or their statistical relevance. This approach accounts not only for how close the features are in value but also for how typical such a match is among known failure cases.

In Equation (5), the traditional similarity measure is adjusted by the coefficient $P_i(A_i, B_i)$, which acts as a credibility filter. Even when values A_i and B_i are numerically close, the final similarity score is lowered if their match is statistically rare in diagnostic practice. Conversely, moderate similarity may be weighted more heavily if it frequently appears in similar failure scenarios.

For example, if the “oil temperature” differs by only 2°C between two precedents, but such values occur under different operating conditions, the similarity would be overestimated without probabilistic correction.

The use of the multiplier P_i mitigates this effect. Oil temperature can be measured in various contexts: in the crankcase of a diesel engine – indicating overall thermal state; at the radiator/heat exchanger inlet – reflecting cooling effectiveness and thermal load; after the radiator – critical for assessing the cooling system's condition; in the gearbox – especially relevant for ships using electric propulsion via gearboxes.

Advantages of this approach include: flexibility: it does not alter the logic of the existing model; improved reliability: incorporates real-world probabilities; interpretability: each contribution to the final metric has a statistical justification; ease of integration: can be implemented using an existing case database.

If weight optimization or model training is later applied, the probabilistic multiplier may either be adapted or replaced with a conditional probability derived from Naive Bayes models or decision trees. To evaluate the influence of the probabilistic component on the final similarity score between precedents, a comparative plot (Fig. 2) was constructed, showing the difference between the classical and the probability-weighted similarity metrics.

As shown in Figure 2, the similarity index weighted by probability is often lower, especially for rare failures, because critical features are assigned greater importance. When such features match, however, the resulting similarity score is higher. This leads to a reordering of the ranking and a more accurate risk assessment. The probabilistic approach enhances the model’s sensitivity to rare but significant similarities, which is crucial for diagnosing failures in SPP.

The optimization of parameter weights allows the model to reflect their diagnostic relevance. The coefficients $\alpha_{s.i}$, $\beta_{s.i}$ and $\gamma_{s.i}$ are optimized using the Limited-memory Broyden–Fletcher–Goldfarb–Shanno algorithm with Box constraints (L-BFGS-B) [19, 20], which minimizes the following loss function:

$$J(\alpha_{s.i}, \beta_{s.i}, \gamma_{s.i}) = \sum_i^M (S_{norm}(A, B_i) - y_i)^2,$$

where y_i - binary label (1 if the precedents are considered similar, 0 otherwise) for the training set of M known pairs.

The optimization process included several stages:

- Formation of a training dataset containing precedent pairs with known expert-assigned similarity labels;
- Initialization of the weight coefficients;
- Evaluation of optimization quality on a test dataset.

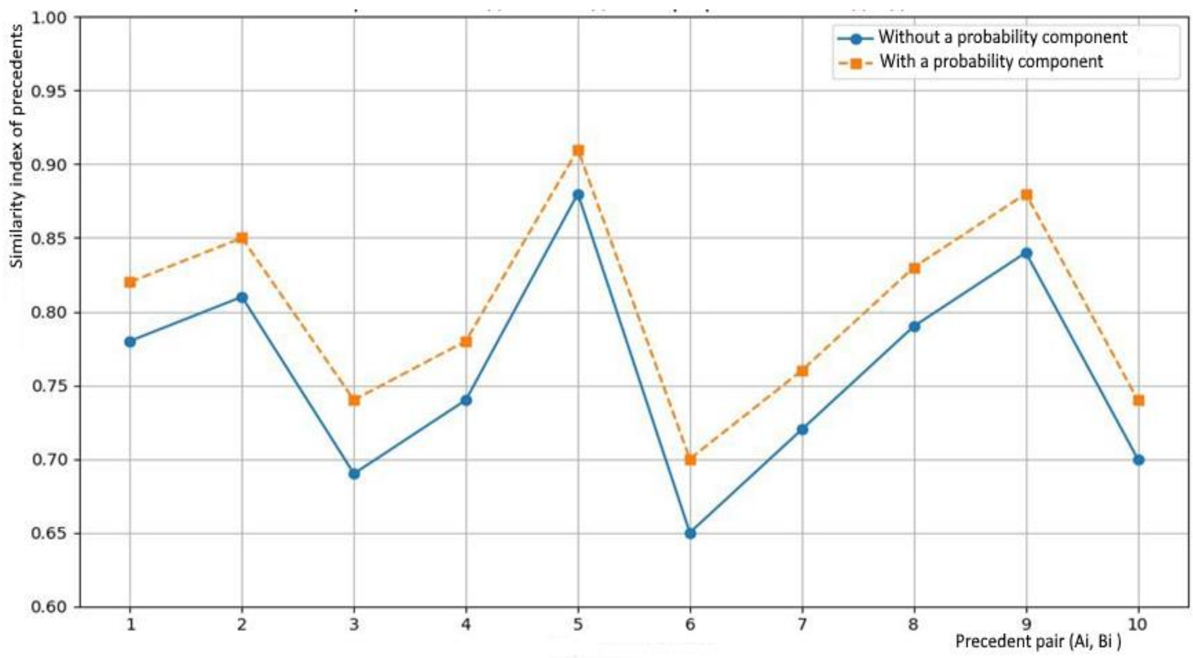

Fig. 2 – Comparison of similarity index between precedents with and without the probabilistic component

The L-BFGS-B method was chosen due to its efficiency under constraint conditions and its good convergence behavior on small datasets. Unlike standard gradient descent, it does not require manual step size selection and inherently respects weight boundaries. The optimization is implemented using scipy.optimize, with specified initial weights, a custom error function, and its minimization.

```
import numpy as np
from scipy.optimize import minimize

# Reference similarity scores (from the training set)
y_true = np.array([...])  # expert evaluations

# Function to compute similarity between precedent pairs based on weights
def similarity_model(weights, data_pairs):
    similarities = []
    for pair in data_pairs:
        # Combination of different similarity metrics (hypothetical example)
        sim = (weights[0] * cosine_similarity(pair[0]['type'], pair[1]['type']) +
               weights[1] * euclidean_similarity(pair[0]['prob'], pair[1]['prob']) +
               weights[2] * jaccard_similarity(pair[0]['risk'], pair[1]['risk']) +
               weights[3] * jaccard_similarity(pair[0]['subsys'], pair[1]['subsys']))
        similarities.append(sim)
    return np.array(similarities)

# Objective function: MSE between expert labels and model predictions
def objective(weights, data_pairs, y_true):
    y_pred = similarity_model(weights, data_pairs)
    return np.mean((y_true - y_pred) ** 2)

# Initial weights
initial_weights = [0.25, 0.25, 0.25, 0.25]
# Bounds: each weight must be between 0 and 1
bounds = [(0, 1)] * 4

# Optimization
result = minimize(objective, initial_weights, args=(data_pairs, y_true), method='L-BFGS-B', bounds=bounds)
optimal_weights = result.x
```

The provided code implements the process of optimizing the weights of parameters used in evaluating the similarity between fault precedents. The optimization is performed using the L-BFGS-B method based on minimizing the mean squared error (MSE) between model predictions and expert assessments of similarity.

Figure 3 illustrates the behavior of the objective function (mean squared error) over the course of the L-BFGS-B optimization iterations.

As shown in Figure 3, the error value rapidly decreases during the initial steps and stabilizes after 10-15 iterations, indicating convergence of the method and the achievement of stable optimal weight values.

To demonstrate the similarity assessment model, two failure precedents (A and B) of SPP equipment are compared across four parameters: failure type, failure probability, risk category, and affected subsystems. Table 2 provides an example of similarity calculation between two precedents, illustrating the application of the model.

The final similarity between precedents A and B is 0.958, indicating a high degree of parametric closeness. The model offers a flexible and interpretable approach to fault similarity assessment, accounting for heterogeneous data types – numerical, categorical, and set-based features. It utilizes specialized similarity metrics, trainable weights based on labeled datasets, and normalization of the final score. This configuration ensures accurate and adaptive diagnostics, enhancing the reliability of identifying analogous cases under operating conditions of SPPs.

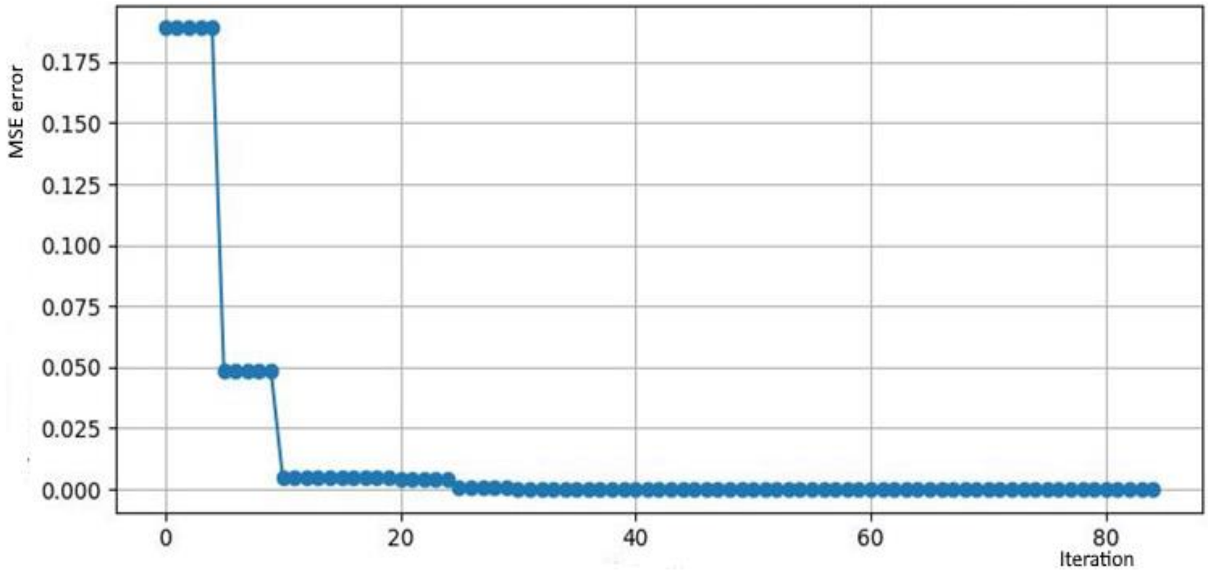

Fig. 3 – Behavior of the objective function (mean squared error) during optimization iterations

Table 2

Example of similarity calculation between two failure precedents

Parameter	Precedent A	Precedent B	Type	Partial Similarity s_i	Weight w_i	Contribution $w_i \cdot s_i$
Equipment Type	Diesel Generator	Diesel Generator	categ.	1.0	0.15	0.15
Operating Time to Failure	1200 h	1000 h	num.	0.83	0.10	0.083
Failure Type	Overheating	Overheating	categ.	1.0	0.20	0.20
Temperature, °C	85	90	num.	0.92	0.15	0.138
Pressure, MPa	2.1	2.0	num.	0.95	0.15	0.143
Vibration, mm/s	4.5	4.8	num.	0.94	0.10	0.094
Operating Conditions	Tropics	Tropics	categ.	1.0	0.15	0.15
					Sum	**0.958**

Conclusions

This paper addresses the development of a unified similarity assessment model for precedents, aimed at application within intelligent decision support systems for diagnosing failures in ship power plant (SPP) equipment. The proposed model accounts for the heterogeneous nature of diagnostic data and implements an adaptive approach to selecting similarity metrics – Euclidean, cosine, and Jaccard – depending on the feature type (numerical, categorical, or set-based). To improve the accuracy of matching, feature weighting is introduced based on diagnostic relevance, implemented using numerical optimization. This ensures the model's flexibility and adaptability to changing operational conditions and equipment configurations.

The formalized structure of a precedent and the proposed classification of typical failures provide a standardized representation of diagnostic information, simplifying the implementation of matching algorithms within the CBR methodology. A review of existing methods revealed their limitations when applied to incomplete, heterogeneous, and poorly structured data, underscoring the relevance of the proposed approach. The matching algorithm, which incorporates operational context, demonstrates interpretability, scalability, and practical applicability in intelligent diagnostic systems. Its deployment improves the reliability and timeliness of failure identification and contributes to the overall reliability, safety, and efficiency of ship power plant operations – particularly under conditions of limited training data and high failure variability.

References

[1] V. Vychuzhanin, and A. Vychuzhanin, *Stochastic Models and Methods for Diagnostics, Assessment, and Prediction of the Technical Condition of Complex Critical Systems*. Lviv-Torun: Liha-Pres, 2025. **doi: 10.36059/978-966-397-457-6**.

[2] N. Neykov, and S. Stefanova, «Using case-based reasoning in system diagnostics and maintenance», in Proc. of 7-th Int. Congress on Information and Communication Technology. Lecture Notes in Networks and Systems, 2023, vol. 464, pp. 305-312. **doi: 10.1007/978-981-19-2394-4_28**.

[3] Z. Chen, K. Wang, C. Zhang, and Q. Wang, «Graph neural network-based fault diagnosis: A review», 2021, *arXiv:2107.04542*. **doi: 10.48550/arXiv.2107.04542**.

[4] H. Lin, Q. Wu, L. Liu, and X. Song, «Case-based reasoning system for aeroengine fault diagnosis enhanced with attitudinal Choquet integral», *Applied Sciences*, vol. 12, no. 3, article 1130, 2022. **doi: 10.3390/app12031130**.

[5] J. Ye, «Single-valued neutrosophic similarity measures based on cotangent function and their application in the fault diagnosis of steam turbine», *Journal of Intelligent & Fuzzy Systems*, vol. 32, no. 6, pp. 4229-4236, 2017. **doi: 10.3233/JIFS-169424**.

[6] J. Serrà, and J.L. Arcos, «An empirical evaluation of similarity measures for time series classification». *Pattern Recognition Letters*, vol. 35, pp. 45-52, 2014. **doi: 10.1016/j.patrec.2012.12.020**.

[7] I. El Bitar, F.Z. Belouadha, and O. Roudies, «A logic and adaptive approach for efficient diagnosis systems using CBR», *International Journal of Computer Applications*, vol. 39, no. 15, pp. 1-5, 2012. **doi: 10.5120/4893-7393**.

[8] V. Vychuzhanin, N. Shibaeva, A. Vychuzhanin, and N. Rudnichenko, «Intellectualization Method and Model of Complex Technical System's Failures Risk Estimation and Prediction», in Proc. of the 6-th Int. Workshop «Computer Modeling and Intelligent Systems», Zaporizhzhia, Ukraine, 2023, vol. 3392, pp. 130-140. **doi: 10.32782/cmis/3392-11**.

[9] V. Vychuzhanin, and N. Rudnichenko, «Complex Technical System Condition Diagnostics and Prediction Computerization», in Proc. of the 3-d Int. Workshop «Computer Modeling and Intelligent Systems», Zaporizhzhia, Ukraine, 2020, vol. 2608, pp. 1-15. **doi: 10.32782/cmis/2608-4**.

[10] H. Wang, and W. Wang, «Bearing fault diagnosis method based on similarity measure and ensemble learning», *Measurement Science and Technology*, vol. 32, no. 5, 2021. **doi: 10.1088/1361-6501/abda97**.

[11] J. Choi, D. Kong, and H. Cho, «Weighted domain adaptation using the graph-structured dataset representation for machinery fault diagnosis under varying operating conditions», *Sensors*, vol. 24, no. 1, article 188, 2024. **doi: 10.3390/s24010188**.

[12] Z. Cheng, and A. Yan, «A case weighted similarity deep measurement method based on a self-attention Siamese neural network», *Industrial Artificial Intelligence*, vol. 1, no. 1, 2023. **doi: 10.1007/s44244-022-00002-y**.

[13] X. Diao, M. Pietrykowski, F. Huang, C. Mutha, and C. Smidts, «An ontology-based fault generation and fault propagation analysis approach for safety-critical computer systems at the design stage», *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, vol. 36, no. 3, pp. 345-360, 2022. **doi: 10.1017/S0890060421000342**.

[14] A. Aamodt, and E. Plaza, «Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches», *AI Communications*, vol. 7, no. 1, pp. 39-59, 1994. **doi: 10.3233/AIC-1994-7104**.

[15] C.D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*. Cambridge: Cambridge University Press, 2008. **doi: 10.1017/CBO9780511809071**.

[16] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. New York: Springer, 2009. **doi: 10.1007/978-0-387-84858-7**.

[17] C.M. Bishop, *Pattern Recognition and Machine Learning*. New York: Springer, 2006.

[18] P.-N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*, 1st ed. Boston: Addison-Wesley, 2005.

[19] P.T. Boggs, and R.H. Byrd, «Adaptive, Limited-Memory BFGS Algorithms for Unconstrained Optimization», *SIAM Journal on Optimization*, vol. 29, no. 2, pp. 947-967, 2019. **doi: 10.1137/16M1065100**.

[20] H. Tankaria, S. Sugimoto, and N. Yamashita, «A Regularized Limited Memory BFGS Method for Large-Scale Unconstrained Optimization and its Efficient Implementations», 2021, *arXiv:2101.04413*. **doi: 10.48550/arXiv.2101.04413**.

МОДЕЛЬ ВИЗНАЧЕННЯ ПОДІБНОСТІ МІЖ ПРЕЦЕДЕНТАМИ ДЛЯ ДІАГНОСТИКИ ВІДМОВ ОБЛАДНАННЯ СУДНОВОЇ ЕНЕРГЕТИЧНОЇ УСТАНОВКИ

Вичужанін В. *д-р техн. наук, професор, Національний університет «Одеська політехніка», м. Одеса, ORCID: https://orcid.org/0000-0002-6302-1832, e-mail: v.v.vychuzhanin@op.edu.ua;*

Вичужанін А. *доктор філософії (Комп'ютерні науки), доцент, Національний університет «Одеська політехніка», м. Одеса, ORCID: https://orcid.org/0000-0001-8779-2503, e-mail: vychuzhanin.o.v@op.edu.ua*

Сучасні суднові енергетичні установки (СЕУ) є високотехнологічними комплексами, функціонування яких супроводжується постійним збором та аналізом великих обсягів діагностичних даних. Підвищення ефективності технічної експлуатації таких систем потребує розробки інтелектуальних засобів діагностики, здатних до точного розпізнавання й інтерпретації типових відмов. У статті розглянуто задачу оцінювання ступеня подібності між прецедентами в рамках методології Case-Based Reasoning (CBR), що є актуальним підходом за умов обмеженої навчальної вибірки та високої варіативності відмов. Подано аналітичний огляд сучасних методів оцінювання подібності, зокрема метричних підходів, агрегуючих операторів (зокрема інтеграла Чокета), логіко-вивідних систем та семантичних графових моделей. Визначено основні обмеження цих підходів: низька адаптивність до гетерогенних даних, слабка інтерпретованість результатів і висока залежність від апріорної інформації. Авторами запропоновано уніфіковану модель визначення подібності, яка враховує різнорідність діагностичних ознак. У моделі реалізовано інтеграцію різних метрик (евклідової, косинусної та Жаккара) залежно від типу даних (числові, категоризовані, множинні). Діагностична значущість ознак задається ваговими коефіцієнтами, які визначаються чисельною оптимізацією. У роботі також подано формалізований опис структури прецедента, класифікацію типових відмов та алгоритм зіставлення з урахуванням контексту експлуатації. Запропонований підхід забезпечує інтерпретованість, адаптивність і масштабованість моделі, що підтверджує її застосовність в інтелектуальних системах підтримки прийняття рішень для діагностики відмов обладнання СЕУ. Практична реалізація моделі сприятиме підвищенню надійності, безпеки й ефективності експлуатації морських технічних систем.

Ключові слова: *інтелектуальні системи підтримки прийняття рішень, суднова енергетична установка, відмова обладнання, прецедент, Case-Based Reasoning, подібність прецедентів, метрика, інтерпретованість.*

Перелік використаних джерел

[1] Vychuzhanin V., Vychuzhanin A. Stochastic Models and Methods for Diagnostics, Assessment, and Prediction of the Technical Condition of Complex Critical Systems. Lviv-Torun: Liha-Pres, 2025. 176 p. DOI: https://doi.org/10.36059/978-966-397-457-6.

[2] Neykov N., Stefanova S. Using case-based reasoning in system diagnostics and maintenance. *Proceedings of Seventh International Congress on Information and Communication Technology. Lecture Notes in Networks and Systems*. 2023. Vol. 464. Pp. 305-312. DOI: https://doi.org/10.1007/978-981-19-2394-4_28.

[3] Graph neural network-based fault diagnosis: A review / Chen Z., Wang K., Zhang C., Wang Q. arXiv preprint. 2021. arXiv:2107.04542. DOI: https://doi.org/10.48550/arXiv.2107.04542.

[4] Case-based reasoning system for aeroengine fault diagnosis enhanced with attitudinal Choquet integral / Lin H., Wu Q., Liu L., Song X. *Applied Sciences*. 2022. Vol. 12, no. 3. Article 1130. DOI: https://doi.org/10.3390/app12031130.

[5] Ye J. Single-valued neutrosophic similarity measures based on cotangent function and their application in the fault diagnosis of steam turbine. *Journal of Intelligent & Fuzzy Systems*. 2017. Vol. 32, no. 6. Pp. 4229-4236. DOI: https://doi.org/10.3233/JIFS-169424.

[6] Serrà J., Arcos J.L. An empirical evaluation of similarity measures for time series classification. *Pattern Recognition Letters*. 2014. Vol. 35. Pp. 45-52. DOI: https://doi.org/10.1016/j.patrec.2012.12.020.

[7] El Bitar I., Belouadha F.Z., Roudies O. A logic and adaptive approach for efficient diagnosis systems using CBR. *International Journal of Computer*

Applications. 2012. Vol. 39, no. 15. Pp. 1-5. DOI: https://doi.org/10.5120/4893-7393.

[8] Intellectualization Method and Model of Complex Technical System's Failures Risk Estimation and Prediction / Vychuzhanin V., Shibaeva N., Vychuzhanin A., Rudnichenko N. *Computer Modeling and Intelligent Systems* : Proceedings of The Sixth International Workshop, Zaporizhzhia, Ukraine, 3 May 2023. Vol. 3392. Pp. 130-140. DOI: https://doi.org/10.32782/cmis/3392-11.

[9] Vychuzhanin V., Rudnichenko N. Complex Technical System Condition Diagnostics and Prediction Computerization. *Computer Modeling and Intelligent Systems* : Proceedings of The Third International Workshop, Zaporizhzhia, Ukraine, 27 April -1 May 2020. Vol. 2608. Pp. 1–15. DOI: https://doi.org/10.32782/cmis/2608-4.

[10] Wang H., Wang W. Bearing fault diagnosis method based on similarity measure and ensemble learning. *Measurement Science and Technology*. 2021. Vol. 32, no. 5. DOI: https://doi.org/10.1088/1361-6501/abda97.

[11] Choi J., Kong D., Cho H. Weighted domain adaptation using the graph-structured dataset representation for machinery fault diagnosis under varying operating conditions. *Sensors*. 2024. Vol. 24, no. 1. Article 188. DOI: https://doi.org/10.3390/s24010188.

[12] Cheng Z., Yan A. A case weighted similarity deep measurement method based on a self-attention Siamese neural network. *Industrial Artificial Intelligence*. 2023. Vol. 1, no. 1. DOI: https://doi.org/10.1007/s44244-022-00002-y.

[13] An ontology-based fault generation and fault propagation analysis approach for safety-critical computer systems at the design stage / X. Diao et al. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*. 2022. Vol. 36, no. 3. Pp. 345-360. DOI: https://doi.org/10.1017/S0890060421000342.

[14] Aamodt A., Plaza E. Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches. *AI Communications*. 1994. Vol. 7, no. 1. Pp. 39-59. DOI: https://doi.org/10.3233/AIC-1994-7104.

[15] Manning C.D., Raghavan P., Schütze H. Introduction to Information Retrieval. Cambridge: Cambridge University Press, 2008. 506 p. DOI: https://doi.org/10.1017/CBO9780511809071.

[16] Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd ed. New York: Springer, 2009. 745 p. DOI: https://doi.org/10.1007/978-0-387-84858-7.

[17] Bishop C. M. Pattern Recognition and Machine Learning. New York: Springer, 2006. 778 p.

[18] Tan P.-N., Steinbach M., Kumar V. Introduction to Data Mining. 1st ed. Boston: Addison-Wesley, 2005. 780 p.

[19] Boggs P.T., Byrd R.H. Adaptive, Limited-Memory BFGS Algorithms for Unconstrained Optimization. *SIAM Journal on Optimization*. 2019. Vol. 29, no. 2. Pp. 947-967. DOI: https://doi.org/10.1137/16M1065100.

[20] Tankaria H., Sugimoto S., Yamashita N. A Regularized Limited Memory BFGS Method for Large-Scale Unconstrained Optimization and its Efficient Implementations. arXiv preprint. 2021. arXiv:2101.04413. DOI: https://doi.org/10.48550/arXiv.2101.04413.

Стаття надійшла 23.04.2025
Стаття прийнята 15.05.2025
Стаття опублікована 30.06.2025

Цитуйте цю статтю як: Vychuzhanin V., Vychuzhanin A. Model for determining similarity between cases in failure diagnosis of ship power plant equipment. *Вісник Приазовського державного технічного університету. Серія: Технічні науки*. 2025. Вип. 50. С. 9-17. DOI: https://doi.org/10.31498/2225-6733.50.2025.336233.