

Data Asset Management and Visualization Based on Intelligent Algorithm: Taking Power Equipment Data as An Example

Xiaoming XIE*, Yixuan PENG

Abstract: Data asset management is adequate in solving the problem of data silence and data idleness for enterprises. Through intelligent algorithms such as neural network, in-depth learning and block chain, and guided by business needs, it extracts, analyzes and visualizes the existing business precipitation data, and forms scattered and disordered data into valuable information to support the development of the company, so as to activate data assets. Taking the management data of electric power equipment as an example, this paper proposes a method of fusion of multiple intelligent control algorithms. The specific modules include the fusion of heterogeneous data; feature extraction of equipment asset management data based on machine learning; intelligent control of multi-objective optimization environment based on energy consumption data; BIM data visualization based on data classification-energy extraction-neural network (SVM-CART-SAE-DNN) algorithm fusion. The algorithm can effectively improve the efficiency of equipment management and enhance the security and economy of power infrastructure through intelligent control of equipment management.

Keywords: data assets; fusion; intelligent algorithms; management; power equipment data; visualization

1 INTRODUCTION

As an intangible asset, data is an important strategic resource for enterprises. For now, the information system construction process of State Grid Corporation of China mainly adopts the "chimney" architecture, which has high construction cost, long cycle, non-sharing of data, low quality, a large number of data deposited in the background of the system, without in-depth mining and application, on the one hand, it occupies tremendous system resources. On the other hand, it does not provide useful decision support for enterprise management, resulting in great waste. Taking the inspection data of power space equipment as an example, this paper introduces the method of data asset extraction, analysis, and display based on artificial intelligence method to aggregate, reconstruct and unify the supervision of data.

As a comprehensive infrastructure, the power space equipment system has many internal risk factors and complex risk types. Once an accident occurs, there will be a significant impact on urban public safety. The health management of power space equipment in China still stays at the stage of regular inspection by appointed personnel, which cannot predict the risk in advance, let alone prevent it from happening. Prognostics and health management (PHM) based on data, as a key method to ensure the safety and reliability of equipment, has achieved fruitful theoretical results and been widely used in practice in the past decades. However, the health management of power space equipment in China is still in its infancy. Problems such as small data dimension and immature algorithm research need to be solved urgently. The management of equipment in power space only relies on a series of manual operations such as staff patrol inspection, fault detection and repair. The operation and maintenance of simple project-level and company-level power space can still operate normally, but the following serious drawbacks are exposed under the city-level power space supervision system: Firstly, in the city-level super-large-scale power space environment, manual inspection requires huge manpower and material resources, and it is difficult for managers to dispatch materials and personnel. Second,

inspectors can only find problems when the equipment fails.

The operation management research on electric power space equipment started earlier in the USA. On the basis of collecting a large number of data, Hiromitsu et al. analyzed and determined the hazards and risk categories existing in electric power space, and put forward preventive measures and methods for different hazards and risks [2]. Farhad et al. classified and summarized that power space disasters mainly include fire, power outage, explosion, etc., and designed different countermeasures and preventive methods for different types of disasters [3]. The environmental and psychological problems are further studied, and the importance of safety monitoring and safety management in the power space is put forward. The research method on the operational management of power space security risk includes automatic control etc. [4-5]. Van et al. studied the security of power space. The existing problems are summarized, including economic problems, development problems, cost problems, environmental problems, regulatory issues, etc., and the lack of successful power space safety management experience is highlighted [6]. According to Klepikov's research, human health, psychology, safety, and other factors should be considered comprehensively to improve the utilization rate of power space equipment [7]. Yoo et al. developed a tunnel risk assessment system based on information technology (IT).

With analyzing the potential risks of the tunnel by GIS and AI, the system is developed in the geographic information system (GIS) environment [8]. Canto-Perello further studied the risk assessment of personnel entering the system and analyzed the potential risks. The result is that accessibility and maintainability are the key points which distinguish power space from other public facilities [9]. Rogers studied the air pollutant diffusion standard and the design and development of space ventilation systems [10].

With regard to the conflagration in power space, some scholars carried out research on Fire Characteristics of Urban Power Space Based on CFD Simulation Analysis. Under the condition of setting fire scale, they analyze the influence of different fire compartment lengths and different ventilation velocities on fire characteristics in

detail [11]. On this basis, the optimal design length of fire compartment and ventilation velocity for smoke control of such facilities are proposed [12]. Zhang analyzed the types and characteristics of fire risk in power space and proposed practical measures to reduce fire risk in power space [13]. Yue made a systematic study of urban power space with the aspects of seismic response analysis and seismic reliability [14].

2 RESEARCH METHOD

The existing research on the safety risk of urban power space equipment mostly focuses on the accident risk of single equipment. It fails to predict the composite and coupling risk of power space and needs to consider the impact of multiple factors on the risk of power space. With the development of Internet of Things (IoT), collecting and analyzing data outside the power space is possible. With the perspective of equipment failure rate (EFR) feature extraction and multi-objective optimization environment intelligent regulation and control, this paper focuses on a set of power space environment intelligent regulation and control system based on EFR. The detail is shown in Fig. 1.

Since support vector machines and neural networks are black-box models, features cannot be selected directly during model learning, while decision tree models have some feature selection capability in model learning. In this paper, the decision trees are combined with support vector machines and neural networks. By building a decision tree, considering the importance of features, extracting features using the height of the tree nodes, and feeding the features with higher classification ability into the support vector machine and neural network models for training, the factors affecting the equipment lifetime and the mechanism of influence are extracted and analyzed. Multi-objective planning is a branch of mathematical planning. Multiple objective functions are considered to be optimized over a given domain. It is often called MOP (multi-objective planning). The planning decision-maker can propose an expected value (or satisfaction value) for each objective function; the solution to the problem is selected by comparing the deviation between the actual and expected values. In this paper, we extract the characteristics of

factors affecting EFR based on environmental data and adjust the energy consumption data by combining various environmental indicators.

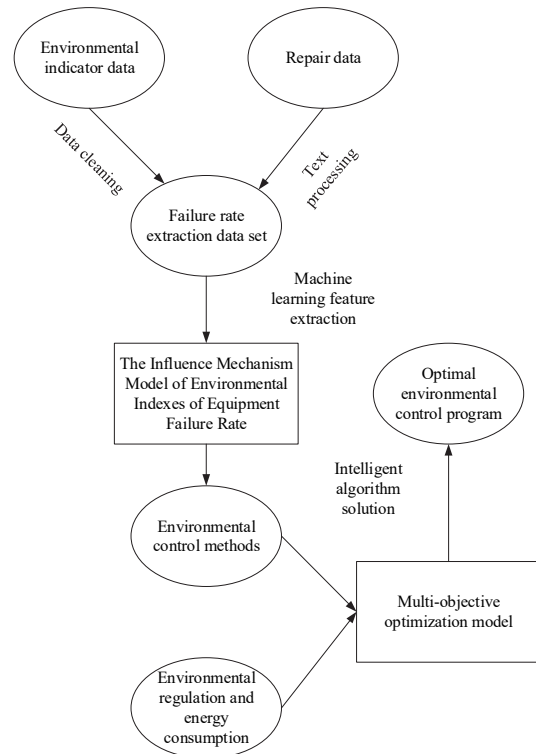


Figure 1 Power space environment intelligent control system based on EFR

3 PRE-PROCESSING OF POWER SPATIAL DATA

3.1 Data Description

The power space is exposed to complex accident risks, including high temperature, high humidity, water accumulation, smoke, and other unfavorable conditions. Therefore, it is required to collect the real-time data of vibration, acoustic and optical signals, temperature, humidity, water level, toxic and hazardous gas concentration, oxygen concentration, smoke particle concentration, and inspection and maintenance by maintenance personnel in the tube. The environment data is shown in Tab. 1.

Table 1 Space environment data

CREATE_TIME	TEMPERATURE	HUMIDITY	O2	CH4	H2S	WATER_WARN
2019-4-22 11:39	10	91	21	0	0	0
2019-4-22 11:40	10	91	21	0	1	0
2019-4-22 11:40	9	80	21	0	0	0
2019-4-22 11:40	9	91	21	0	0	0
2019-4-22 11:40	11	94	21	0	0	0
2019-4-22 11:40	8	79	21	0	0	0
2019-4-22 14:43	10	86	21	9	0	0
2019-4-22 14:43	10	86	21	9	0	0
2019-4-22 14:44	10	86	21	9	0	0
2019-4-22 14:45	9	89	21	6	0	0
2019-4-22 14:46	9	89	21	6	0	0

Event	ID	Name	Time
5bb418e3554b4e97bc32d437680	09010106-RL-013	HUM_OVERLIMIT	2019/4/21 10:07
66a2cf45ae1b4e9799550467e71e	09010106-T-003	TMP_OVERLIMIT	2019/4/21 10:15
7b3ee96a42864cdfadc26e2c744d	09010106-Z-004	HUM_OVERLIMIT	2019/4/21 10:23
020d403e4ad54f69bfd938a4282b	09010106-RL-013	HUM_OVERLIMIT	2019/4/21 10:28
0a466c4534544a49b77e8777c1e	09010108-D-005	O2_OVERLIMIT	2019/4/21 10:43
277485521de14e53b686b8bb80c	09010106-Z-004	HUM_OVERLIMIT	2019/4/21 10:44
05dd244b43d745549aeb13b6d34	09010106-T-003	TMP_OVERLIMIT	2019/4/21 10:46
9b864060949548bd996259c7fb00	09010106-T-006	HUM_OVERLIMIT	2019/4/21 10:46
f4e29c1d80e94448acc67e50231f1	09010106-T-006	HUM_OVERLIMIT	2019/4/21 10:00
de214b87f5744c53b95402293511	09010106-T-003	TMP_OVERLIMIT	2019/4/21 10:03
72cbb4935b044c2282b4f2e202a9	09010108-D-005	O2_OVERLIMIT	2019/4/21 10:04
00704fd38e7047c4b3775cc598b0	09010108-D-005	O2_OVERLIMIT	2019/4/21 10:07
65c999bfb6524c5b87f9785ca785e	10030102-SB-034	PRSN_INTRUSION	2019/4/21 10:09
360a045f4cc54ac1a389db90f4509	09010106-S-020	HUM_OVERLIMIT	2019/4/21 10:10
a96370abe25c4460ba6ad2e364f3	09010106-Z-004	HUM_OVERLIMIT	2019/4/21 10:14
e8f10160639544a6bae02368941c	09010106-S-020	HUM_OVERLIMIT	2019/4/21 10:16
291a53f73fa24eef9be4b4501c629	09010108-D-005	O2_OVERLIMIT	2019/4/21 10:16

Figure 2 Warranty data of patrol personnel

ID	DEV_ID	EVT_NAME	ALM_TIME
f45cbb7d88f448eab0b836a840f763aa	09010106-S-020	HUM_OVERLIMIT	2019/4/22 10:24
86ed92b2ef954999e88a68bf6d1c57c	09010106-S-020	HUM_OVERLIMIT	2019/4/22 12:38
a22c75a4d7574f3aa83dff7ca816772	09010106-S-020	HUM_OVERLIMIT	2019/4/22 12:48
7a37ffd22d154cf4b7ce9b66a4f7c04f	09010106-RL-013	HUM_OVERLIMIT	2019/4/22 12:57
3531b1741fec1329fa70de0defa09ec	09010106-S-020	HUM_OVERLIMIT	2019/4/22 12:57
9257bd94c1614adaaa1eb22e240592ba	09010106-S-020	HUM_OVERLIMIT	2019/4/22 13:06
075ec89618aa4126bab255a21e3d63a3	09010106-RL-013	HUM_OVERLIMIT	2019/4/22 13:12
0c6951d4dbb640c6bb6d457b4db6c194	09010106-S-020	HUM_OVERLIMIT	2019/4/22 13:12
1f5b2c1945ab4047a492d05b49f41a97	09010106-RL-013	HUM_OVERLIMIT	2019/4/22 13:15
8efaa6078b274ad6abc9f75a7d110912	09010106-RL-013	HUM_OVERLIMIT	2019/4/22 13:18
8025b7f69356408b9c9609be39722d9d	09010106-S-020	HUM_OVERLIMIT	2019/4/22 13:28
6f70c193a8af406498aef872120b6c6a	09010106-RL-013	HUM_OVERLIMIT	2019/4/22 13:30
e1b3ebbad30b454c8b9e29c4cf2dd2c5	09010106-S-020	HUM_OVERLIMIT	2019/4/22 13:30
3303164bf1e1485086dcd450ccc82136	09010106-S-020	HUM_OVERLIMIT	2019/4/22 13:34
cfe11e0d1cf34e68a8974f1890bb97a1	09010106-RL-013	HUM_OVERLIMIT	2019/4/22 13:43
39a2aca11aa44886b8797f3dabbcb3bd0	09010106-S-020	HUM_OVERLIMIT	2019/4/22 13:55
fb17953dff384c23a354fcb75190819d	09010106-RL-013	HUM_OVERLIMIT	2019/4/22 14:01
6e119f1420d5462695cf3bdc4c2d6e33	09010106-S-020	HUM_OVERLIMIT	2019/4/22 14:01
d1c56b73cbec4566af8593b6c9550c36	09010106-RL-013	HUM_OVERLIMIT	2019/4/22 14:04

Figure 3 Warranty data of staff patrol inspection

The independent variable is a time series panel data of spatial environmental indicators, including equipment failure information, while the equipment failure information in the power space is manually input at this stage, and the content is mostly unstructured text and picture information. In this paper, multi-source data fusion technology based on data pattern matching is applied to preprocess multi-source data (Fig. 2 and Fig. 3). With the rapid development of IoT, the Internet of Everything has become an inevitable trend. IoT involves smart homes, smart transportation etc., enabling people to connect with any person or device anytime, anywhere. However, most of the heterogeneous data in IoT is stored in silo. It hinders the pace of the Internet of Everything. Data pattern matching is widely used in data association, which can solve the above problems.

3.2 Multi-Element Data Fusion

3.2.1 SVM Text Classification

As shown in Fig. 2 and Fig. 3, the SVM text classification algorithm is used to preprocess the numerous and complex employee data. Text classification is a classic topic in natural language processing, and free text files need to be assigned predefined categories. To date, almost all text classification techniques are based on words, where simple statistics of ordered combinations of some words, such as n-grams, usually perform best [15]. SVM can transform the original space into a linear problem of high-dimensional space through nonlinear transformation, and

choosing the appropriate kernel function is crucial to SVM [16].

1. Text feature extraction and representation:

At present, in order to simplify the feature selection process when extracting text features, the assumption of feature independence is often adopted to achieve a compromise between computing time and computing quality. The general method is to select the best features as a subset of text features by setting feature thresholds according to the feature vectors of words in the text, and then build a feature model. Thus, we use the TF-IDF formula to calculate the weight of a word.

$$\omega_{ik} = \frac{tf_{ik} \cdot \log\left(\frac{N}{n_k} + 0.01\right)}{\sqrt{\sum_{i=k} \left[f_{ik} \cdot \log\left(\frac{N}{n_k} + 0.01\right) \right]^2}} \quad (1)$$

where tf_{ik} represents the frequency of the occurrence of the word t_i in the document d_i at the special diagnosis time t_k , N represents the total number of training documents, and n_k represents the number of documents in which the word t_i appears in the training set. Thus, the higher the frequency of a word in a batch of documents, the lower the degree of discrimination and the lower the weight.

2. Normalization processing:

Normalization is to limit the data to a certain range in the process of algorithm. As we can see, a denotes the word frequency of a keyword, \min and \max respectively denote the minimum and maximum word frequency of the word in all texts. Normalization helps to make text classification more accurate.

$$\frac{a - \min}{\max - \min} = b \tag{2}$$

3. Text categorization:

After text preprocessing, feature extraction, feature representation and normalization, the original text information is abstracted into a vectorized sample set, and then the similarity is calculated between the sample set and the trained template file, and if the sample set does not belong to a category, the similarity is calculated with the template file of other categories until the sample is classified into the corresponding category. Fig. 4 depicts how the SVM model classifies the text.

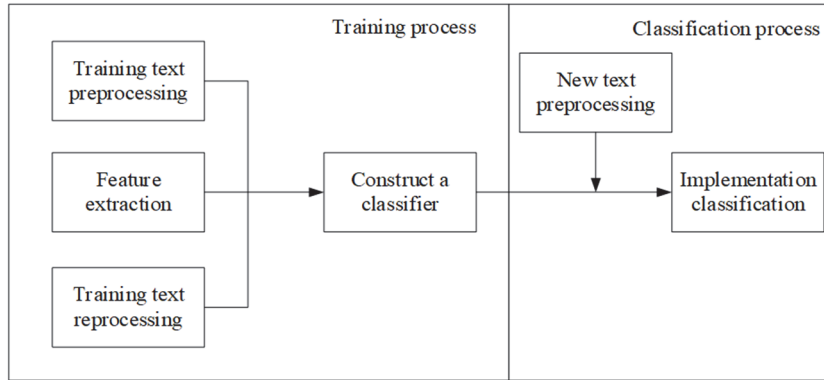


Figure 4 System implementation based on SVM

3.2.2 Pattern Matching Based on Cabin and Time

Pattern matching algorithms measure the similarity of pairs of elements from different data patterns in a feature and output a real value in the interval [0, 1] as the similarity of two input elements from different data patterns. After investigation, this paper summarizes the classification of

basic pattern matching algorithms [17]. As shown in Fig. 5, the processing of element pairs in two input heterogeneous data patterns consists of three main stages, including similarity calculation phase, similarity synthesis phase, similarity determination phase.

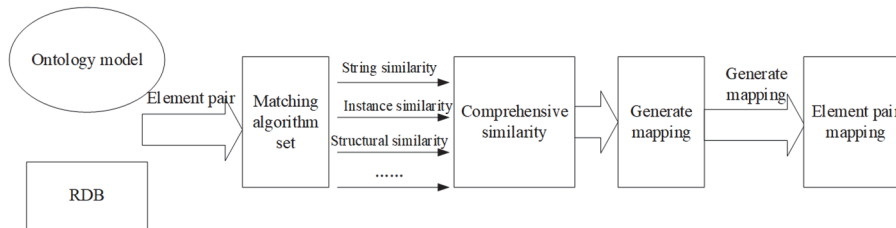


Figure 5 General flow of the pattern matching framework

4 FEATURE EXTRACTION FOR EFR

4.1 CART Model

CART is Classification and Regression Tree. CART assumes that the decision tree is a binary tree, and only "yes" and "no" are allowed in the values of the internal node characteristics. The left branch is the branch with the "yes" value, and the opposite is true if there is a branch. This decision tree is equivalent to recursively bisecting each feature [18-20]. With k classes, the Gini coefficient of the probability distribution and the corresponding binary classification problem are defined below.

$$G(p) = \sum_{k=1}^K P_k (1 - P_k) \tag{3}$$

$$G(p) = 2p(1 - p) \tag{4}$$

In the decision tree problem, the principle of feature selection using Gini coefficient is as follows:

$$G(D, A) = \frac{D_1}{D} G(D_1) + \frac{D_2}{D} G(D_2) \tag{5}$$

The minimum binary regression tree generation algorithm for regression tree generation is as follows:

(1) Select the optimal segmentation variable j and the segmentation point s , traverse the variable j , scan the segmentation point s for the fixed segmentation variable j , and select the pair (j, s) that minimizes the above equation. Where R_m is the divided input space, and c_m is the output value corresponding to the space R_m .

$$\min_{j,s} \left[\min_{c_1} \sum_{x_i \in R_1(j,s)} (y_i - c_1)^2 + \min_{c_2} \sum_{x_i \in R_2(j,s)} (y_i - c_2)^2 \right] \tag{6}$$

(2) Divide the area with the selected pair (j, s) and determine the corresponding output value.

$$R_1(j, s) = \{x | x^{(j)} \leq s\}, R_2(j, s) = \{x | x^{(j)} > s\} \quad (7)$$

$$\bar{c}_m = \frac{1}{N_m} \sum_{x_i \in R_1(j, s)} y_i, x \in R_m, m=1, 2 \quad (8)$$

(3) Continue to invoke step 1 for both subareas until the stop condition. Divide the input space into M regions R_1, R_2, \dots, R_m generates a decision tree.

$$f(x) = \sum_{m=1}^M \bar{c}_m I(x \in R) \quad (9)$$

With k classes, the Gini coefficient of the probability distribution is defined as:

$$G(p) = \sum_{k=1}^K P_k(1 - P_k) \quad (10)$$

The probability distribution Gini coefficient of the binary classification problem is:

$$G(p) = 2p(1 - p) \quad (11)$$

In the decision tree problem, the principle of feature selection using Gini coefficient is as follows:

$$G(D, A) = \frac{D_1}{D} G(D_1) + \frac{D_2}{D} G(D_2) \quad (12)$$

4.2 Influencing Factors Analysis

This phase describes the specification of the tree structure [21-23]. The maximum tree depth is set to 3 levels and the saliency values for split nodes and merge categories are set to 0.05. The minimum number of cases for the parent node and the child nodes is set to 100 and 50 respectively.

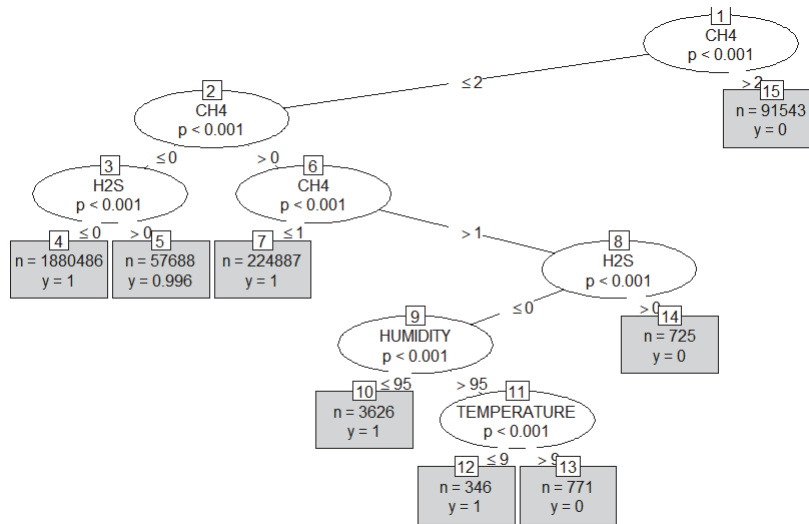


Figure 6 Diagram of CART regression tree model

The results of the model are shown in Fig. 6 for analyzing the relationship between the analysis of equipment life impact factors and the predictive/categorical variables. The final tree structure for equipment life involves five split variables, including temperature, humidity, methane, oxygen and hydrogen sulfide. The first best split for node 0 is based on methane content, which is likely to cause a sharp increase in EFR if it exceeds 2, followed by hydrogen sulfide, which has a greater impact on EFR, with about 1/3 of equipment failures being related to hydrogen sulfide. In addition, humidity and temperature also have some influence on the EFR, but they are not the main factors. When the humidity is too high and the temperature is too high, it is more likely to cause equipment failure.

5 EQUIPMENT FAULT PREDICTION BASED ON NEURAL NETWORK

5.1 Model Establishment

The intelligent regulation of the power space environment is a highly complex nonlinear system, so its design optimization. In order to simplify the problem

analysis and grasp the main factors affecting the problem, the optimization model of intelligent regulation and control system of power space environment is constructed based on the following assumptions: (1) The impact of events on the power space environment environmental indicators is instantaneous. (2) The impact of environmental indicators on EFR does not change with time. (3) The impact of environmental controls on environmental indicators does not vary with time, and the impact of multiple environmental controls on the environment is independent of each other. (4) The total energy consumption is the sum of the energy consumption of each environmental control.

The failure rate of power space control equipment n is affected to some extent by the environmental indicators, such as turning on the fan to reduce humidity. w_{ij} is the duration of initiating loop control i . Generally more than one environmental control operation is required to control an environmental indicator at the same time. In order to make the total energy consumption y_1 and the failure minimization equipment rate y_2 , we start environmental control i for 01 variable x_i and environmental control i for time T_i . Fig. 7 shows the model of intelligent regulation system of power space environment.

The stack-based self-coding (SAE-DNN) training and prediction process consists of three phases. In the first stage, we use a purely unsupervised approach to pre-train the SAE of the greedy layer by combining different features into different training datasets. The SAE is used to initialize the weights and biases of the DNN. In the second stage, the labeled data are used to supervise the DNN final stage prediction. In this paper, a deep neural network based on the SAE-DNN model is built to identify suspicious anomalies [23]. As shown in Fig. 8, feature vectors can be extracted and the trained weights are fed into the DNN, which improves the accuracy of suspicious anomaly detection. Eq. (13) is used to minimize the reconstruction error between the input vector X and the reconstruction vector R .

$$L(X, R) = \min L(X, R) \tag{13}$$

In order to match the reconstruction result R with the input vector X , minimize the loss function $L(X, R)$ by fine-tuning the parameters w_i and b_i .

$$w_i' = w_i - \eta \frac{\partial L(x_i, r_i)}{\partial w_i} \tag{14}$$

$$b_i' = b_i - \eta \frac{\partial L(x_i, r_i)}{\partial b_i} \tag{15}$$

where w_i and w_i' are the original weight and the updated weight in each i node of the hidden layers; b_i and b_i' are the original bias and the updated bias in each i node of the hidden layers; η is the learning rate.

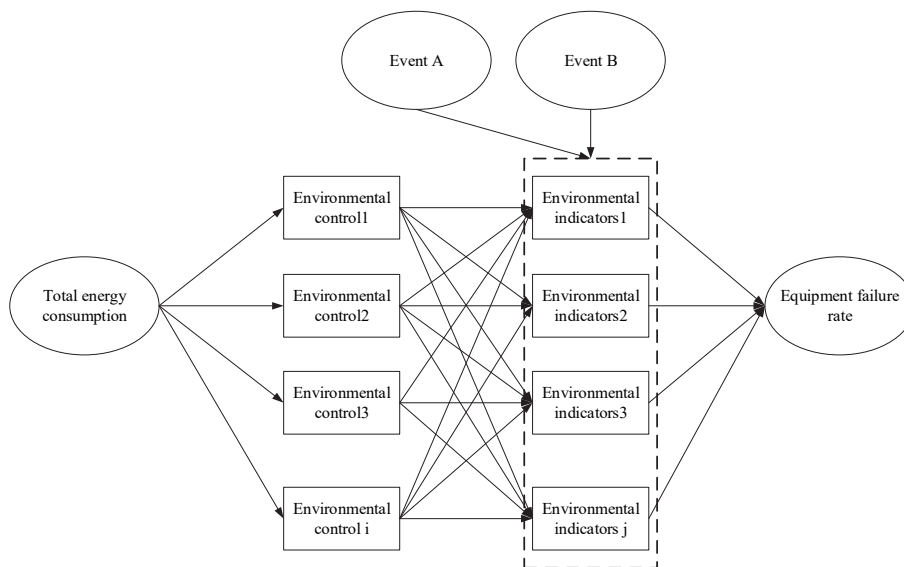


Figure 7 Model of intelligent control system for power space environment

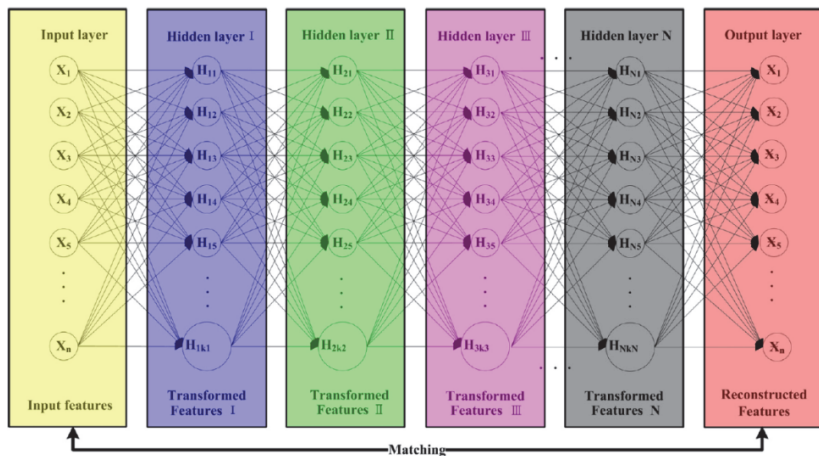


Figure 8 Block diagram of SAE

5.2 Experimental Results

The evaluation indicators are as follows. Where y_i is the real value of the sample, and \bar{y}_i is the corresponding predicted value; T is the number of all samples.

$$RMSE = \sqrt{\frac{1}{T} \sum_{i=1}^T (\bar{y}_i - y_i)^2} \tag{16}$$

$$RMSE = \frac{1}{T} \sum_{i=1}^T \frac{(y_i - \bar{y}_i)}{y_i} \tag{17}$$

Tab. 2 compares the results of the method in this paper with those of the classical machine learning method. It can be seen that the method in this paper is obviously superior to the conventional machine learning method. The RMSE of the final result is 50.69, and the MAPE is 0.1004, which has better fault diagnosis performance.

Tab. 3 compares the prediction method of application environment and operation data with the prediction results of single application environment and operation data, and the results show that the combined application of the two

can obtain better prediction results. Tab. 4 shows five sets of test results.

Table 2 Comparison of prediction effect

Method	RMSE	MAPE
Linear regression	72.94	0.1762
Random forest	56.08	0.1400
Support Vector Machine Regression	54.57	0.1321
Deep Neural Network	55.58	0.1270
Method of this paper	50.69	0.1004

Table 3 Comparison of data types

Data	RMSE	MAPE
Operation Data	58.78	0.1398
Environmental Data	54.29	0.1222
Operation + environment data	50.69	0.1004

Table 4 Test results of a device

Group No.	Environmental Data						Operation Data			Fault prediction error	
	Temperature	Humidity	Oxygen	Methane gas	Hydrogen Sulfide	Water Level	Operating Data 1	Operational Data 2	Operating Data 3	RMSE	MAPE
1	10	91	21	0	0	0				55.58	0.1270
2	10	91	21	0	1	0				50.69	0.1004
3	9	80	21	0	0	0				48.02	0.0885
4	9	91	21	0	0	0				49.91	0.1031
5	11	94	21	0	0	0				52.05	0.1125

5.3 Algorithm Performance Analysis of Solution

The statistical analysis of the Pareto frontier convergence and the average population diversity of the continuous swarm intelligence algorithm selected in the RERFTRD example for 30 independent experiments, and the running time and number of iterations for 30 independent experiments are shown in Fig. 9. The statistical analysis of SC and ISC of the continuous swarm intelligence algorithm in the RERFTRD example over 30

independent experiments shows that the particle swarm optimization algorithm with quantum behavior achieves better optimization results than other algorithms, while the IMOQPSO algorithm achieves the best optimization results. Second, the multi-phase particle swarm optimization algorithm, IMOMPPSO algorithm also achieves better optimization results than the original algorithm. However, the BBMOPSO algorithm achieved the worst optimization results.

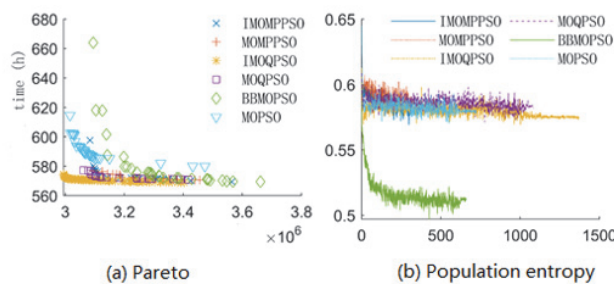


Figure 9 Pareto front convergence and population diversity of continuous swarm intelligence algorithm in RERFTRD example

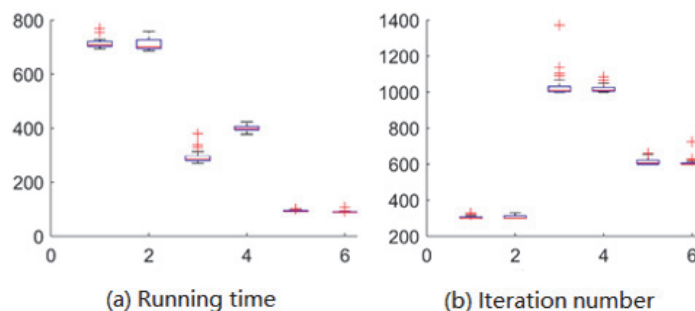


Figure 10 Running time and iteration times of continuous swarm intelligence algorithm in RERFTRD example

6 VISUALIZATION OF DATA ASSETS

The data of this study comes from the power alarm data recorded by the platform, and the original data of the platform is stored in the Oracle database. The time dimension of the alarm data is in seconds, the attributes of the data include the ID of power alarm event (ID), power alarm device ID (DEV_ID), power alarm event name

(EVT_NAME), power alarm time (ALM_TIME), power alarm event level (ALM_LEVEL), and power alarm content (ALM-CONTENT), Operation status (OP_STATE), operation time (OP_TIME), operator (OPERATOR) and ID (ROWID) of the data in the Oracle database. Some examples of raw power alarm data are shown in Fig. 11.

ID	PG_ID	CABIN_ID	REGION_ID	CREATE_TIME	TEMPERATURE	TEM_WARN	HUMIDITY	HUM_WARN	O2	O2_WARN	CH4	CH4_WARN	H2S	H2S_WARN	WATER_WARN	HEALTH	ROWID
1d99f68688SYS_PIPE_GALLSYS_PG_CABIN-(SYS_PG_REGION-				2019/4/22 11:39	10	0	91	0	21	0	0	0	0	0	0	0	100 AAAUPKAAKAADo+LAA
ab37502c6c8SYS_PIPE_GALLSYS_PG_CABIN-(SYS_PG_REGION-				2019/4/22 11:40	10	0	91	0	21	0	0	0	1	0	0	0	100 AAAUPKAAKAADo+LAA
24f6941ca7SYS_PIPE_GALLSYS_PG_CABIN-(SYS_PG_REGION-				2019/4/22 11:40	9	0	80	0	21	0	0	0	0	0	0	0	100 AAAUPKAAKAADo+LAA
504eb7997fSYS_PIPE_GALLSYS_PG_CABIN-(SYS_PG_REGION-				2019/4/22 11:40	9	0	91	0	21	0	0	0	0	0	0	0	100 AAAUPKAAKAADo+LAA
890484d3a1SYS_PIPE_GALLSYS_PG_CABIN-(SYS_PG_REGION-				2019/4/22 11:40	11	0	94	0	21	0	0	0	0	0	0	0	100 AAAUPKAAKAADo+LAA
50a747531fSYS_PIPE_GALLSYS_PG_CABIN-(SYS_PG_REGION-				2019/4/22 11:40	8	0	79	0	21	0	0	0	0	0	0	0	100 AAAUPKAAKAADo+LAA
19da5fc7b4SYS_PIPE_GALLSYS_PG_CABIN-(SYS_PG_REGION-				2019/4/22 11:40	8	0	82	0	21	0	0	0	0	0	0	0	100 AAAUPKAAKAADo+LAA
87eb5b6165SYS_PIPE_GALLSYS_PG_CABIN-(SYS_PG_REGION-				2019/4/22 11:40	9	0	80	0	21	0	0	0	0	0	0	0	100 AAAUPKAAKAADo+LAA
dc0e9a6d6eSYS_PIPE_GALLSYS_PG_CABIN-(SYS_PG_REGION-				2019/4/22 11:40	10	0	83	0	21	0	0	0	0	0	0	0	100 AAAUPKAAKAADo+LAA
1278575d4fSYS_PIPE_GALLSYS_PG_CABIN-(SYS_PG_REGION-				2019/4/22 11:40	8	0	83	0	21	0	0	0	0	0	0	0	100 AAAUPKAAKAADo+LAA
fcaad96d44SYS_PIPE_GALLSYS_PG_CABIN-(SYS_PG_REGION-				2019/4/22 11:40	8	0	82	0	20	0	0	0	0	0	0	0	100 AAAUPKAAKAADo+LAA
954c9395c6SYS_PIPE_GALLSYS_PG_CABIN-(SYS_PG_REGION-				2019/4/22 11:40	8	0	80	0	20	0	0	0	0	0	0	0	100 AAAUPKAAKAADo+LAA
2fc4787218SYS_PIPE_GALLSYS_PG_CABIN-(SYS_PG_REGION-				2019/4/22 11:40	9	0	83	0	21	0	0	0	0	0	0	0	100 AAAUPKAAKAADo+LAA
19f9fd48f65eSYS_PIPE_GALLSYS_PG_CABIN-(SYS_PG_REGION-				2019/4/22 11:40	8	0	83	0	21	0	0	0	0	0	0	0	100 AAAUPKAAKAADo+LAA
16dcfd9396eSYS_PIPE_GALLSYS_PG_CABIN-(SYS_PG_REGION-				2019/4/22 11:40	10	0	84	0	21	0	0	0	0	0	0	0	100 AAAUPKAAKAADo+LAA
850aa43ca1SYS_PIPE_GALLSYS_PG_CABIN-(SYS_PG_REGION-				2019/4/22 11:40	8	0	82	0	20	0	0	0	0	0	0	0	100 AAAUPKAAKAADo+LAA
09a59cccaafSYS_PIPE_GALLSYS_PG_CABIN-(SYS_PG_REGION-				2019/4/22 11:40	9	0	91	0	21	0	0	0	0	0	0	0	100 AAAUPKAAKAADo+LAA
03d7a0f621SYS_PIPE_GALLSYS_PG_CABIN-(SYS_PG_REGION-				2019/4/22 11:40	11	0	87	0	21	0	0	0	0	0	0	0	100 AAAUPKAAKAADo+LAA
b75b73a01SYS_PIPE_GALLSYS_PG_CABIN-(SYS_PG_REGION-				2019/4/22 11:40	10	0	85	0	21	0	0	0	0	0	0	0	100 AAAUPKAAKAADo+LAA
cd9ca7f18eSYS_PIPE_GALLSYS_PG_CABIN-(SYS_PG_REGION-				2019/4/22 11:40	8	0	82	0	21	0	0	0	0	0	0	0	100 AAAUPKAAKAADo+LAA
fe577a36f5SYS_PIPE_GALLSYS_PG_CABIN-(SYS_PG_REGION-				2019/4/22 11:40	10	0	91	0	21	0	0	0	0	0	0	0	100 AAAUPKAAKAADo+LAA
8a4054ce22SYS_PIPE_GALLSYS_PG_CABIN-(SYS_PG_REGION-				2019/4/22 11:40	11	0	88	0	21	0	0	0	0	0	0	0	100 AAAUPKAAKAADo+LAA

Figure 11 Example of partial raw alarm data

Through the observation of power alarm data, it is found that the original data is not complete and there are missing values. Whether missing values can be handled properly or not plays a vital role in data asset management. If the data containing missing values are deleted directly, the judgment of the probability of the occurrence of electric power risk accidents and the judgment of the size of adverse effects once the risk occurs will inevitably have a great impact, the results will be very different from the actual results, and it will also cause a negative impact on the follow-up risk warning and equipment failure prevention work, thus causing the whole risk management work cannot achieve the purpose of risk avoidance or risk reduction. Once the risk evolves into an accident, it will have an irreversible impact on human, material and financial resources. So, it is very important to deal with the missing values in the original data reasonably and apply appropriate and accurate methods to fill in the missing values.

Firstly, the power data is preprocessed, and three useful attributes for the study of equipment fault metrics are screened and separated, which are power alarm event name (EVT_NAME), power pipeline name (LINE_NAME) and power alarm event level (ALM_LEVEL). The power line name is a column separated from the power alarm content (ALM_CONTENT). Power alarm event level is divided into level 1, 2 and 3; there are ten kinds of power pipeline names, namely, entrance and exit (CR), thermal power in the park (RL), gas in the park (RQ), water, information and electricity in the Park (S), equipment room in the Park (SB), natural gas in Baikang Road (T), comprehensive Baikang Road (Z), electric power in Baikang road (D), equipment room in Yankang Road (YSB) and comprehensive Yankang Road (YZ). According to the location of the power park, there are four parts: entrance and exit, park, Baikang Road and Yankang Road. The names of power alarm events are divided into O2_OVERLIMIT, CH4_OVERLIMIT, PRSN_INTRUSION, HUM_OVERLIMIT and TIMEOUT.

An example of partially preprocessed data is shown in Fig. 12:

	A	B	C	D
1	O2 _ OVERLIMIT	RQ		3
2	PRSN _ INTRUSION	RL		fff
3	O2 _ OVERLIMIT	D		t
4	CH4 _ OVERLIMIT	YZ		e
5	PRSN _ INTRUSION	RL		n
6	O2 _ OVERLIMIT	T		t
7	O2 _ OVERLIMIT	S		h
8	?	RQ		o
9	O2 _ OVERLIMIT	YZ		u
10	O2 _ OVERLIMIT	RQ		s
11	O2 _ OVERLIMIT	D		a
12	CH4 _ OVERLIMIT	YZ		d
13	PRSN _ INTRUSION	RL		n
14	PRSN _ INTRUSION	S		o
15	PRSN _ INTRUSION	S		n
16	O2 _ OVERLIMIT	T		e
17	PRSN _ INTRUSION	CR		u
18	PRSN _ INTRUSION	CR		n
19	?	RL		d
20	O2 _ OVERLIMIT	S		r
21	O2 _ OVERLIMIT	RQ		e
22	O2 _ OVERLIMIT	YZ		d
23	PRSN _ INTRUSION	S		a
24	PRSN _ INTRUSION	S		d
25	PRSN _ INTRUSION	S		e
26	PRSN _ INTRUSION	S		le
27	PRSN _ INTRUSION	S		v
28	PRSN _ INTRUSION	S		e
29	PRSN _ INTRUSION	CR		n
30	PRSN _ INTRUSION	CR		1

Figure 12 Example of partially pre-processed data

A typical power failure analysis method consists of four steps, including hazard identification, frequency analysis, consequence analysis and risk quantification. On the basis of collecting the necessary information about power failure to determine the failure mode, risk factors and clear causal relationship, this study obtains the probability of main events and uses dynamic variable weight fuzzy Petri nets to carry out risk reasoning. Initialize the possible result of the event, calculate the relevant consequence probability based on the expected arrival time method, quantify the risk loss using the data classification-fault extraction-neural network algorithm, conduct risk assessment, make risk decisions, and formulate control measures (Fig. 13).

Based on the OnceDI system, a publish/subscribe system, OnceDI/PS, is developed to support data distribution from multiple power data sources. The system consists of two parts: the proxy server and the integrated agent. Its architecture is shown in the figure. The proxy

server system employs a hierarchical, configurable, and scalable architecture.

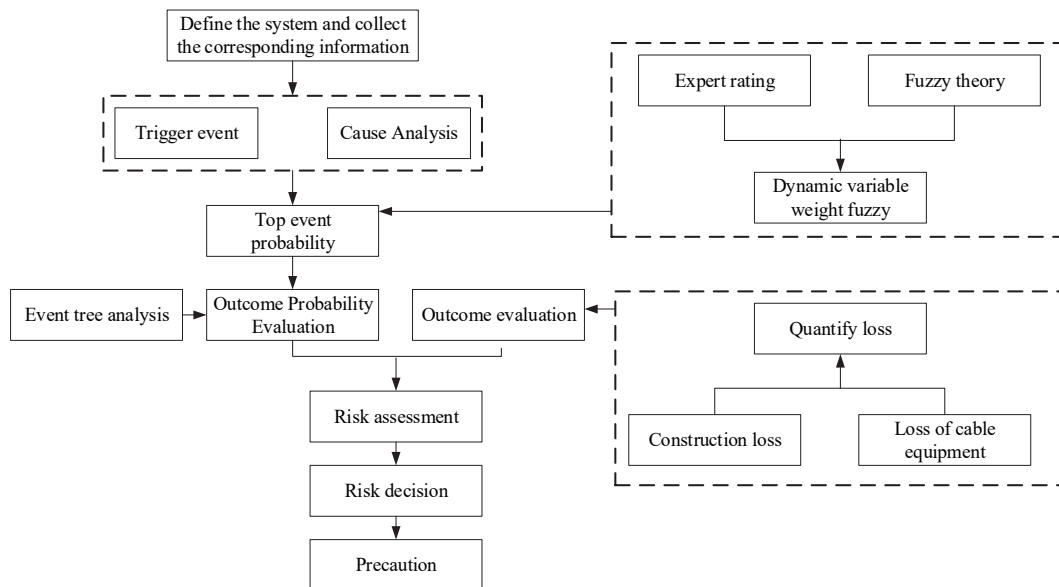


Figure 13 Analysis framework

According to the scalable design requirements of the power publish/subscribe system, the main functions of the proxy server of the power system adopt a hierarchical structure, the system is divided into six layers from top to bottom, including publish/subscribe interface layer, data model and semantic conversion layer, matching layer, routing and scheduling layer, overlay network layer and transport layer. Each layer builds on the functionality provided by the lower layers and provides a clearly defined interface to the upper layers. The layers are independent of each other, and the implementation of each layer is easily replaced by a different implementation.

The publish/subscribe interface layer provides API interfaces for publishing metadata, events, subscriptions, and notifications of the Pub/Sub middleware. The data model and semantic conversion layer are responsible for maintaining the metadata model, semantic event model and semantic subscription model, realizing the semantic conversion of data, managing the metadata and subscription of the system, and maintaining the common vocabulary and mapping function library of the system. The matching layer carries out effective event matching through a matching engine; through the efficient event matching algorithm, all subscriptions satisfied by the event can be found quickly. The routing and scheduling layer is responsible for the forwarding of metadata, subscription and publication data, etc. The overlay network layer is a virtual communication structure logically above the supporting network layer; the main task of this layer is to maintain the topology of the proxy server network, handle the status update information of the overlay network, and deal with the dynamic changes of the overlay network nodes such as joining and leaving. The system provides the realization of acyclic graph and Mesh network. The transport layer is a unicast communication service that represents the underlying network and provides reliable message delivery between proxy servers. In addition, configuration management is the auxiliary facility of the system, which realizes the system configuration of the

server, and configures the transmission protocol, overlay network structure, routing strategy, scheduling algorithm and event matching algorithm according to the application and system requirements.

The proxy server subsystem of the OnceDI/PS system adopts a layered modular design. Standardized interfaces between layers limit the impact of code changes to the layer being changed to only one layer, the layer being changed can be accommodated without changing other layers, independent layer implementations are easily replaced by semantically equivalent implementations. This improves the local dependency, portability, and replaceability of the system.

The BIM-based power equipment operation and maintenance management visualization system is mainly used for the visual display of power equipment information. It is a tool for visual management and can realize the query and statistics of various data of power equipment. In addition, all units involved in operation and maintenance management can access and query relevant information in real time through the platform. The authorized person can also delete, add and update the information in the system. Firstly, BIM software is used for 3D modeling of power equipment, and the basic information, maintenance information, cost information and contract information of the equipment/pipeline are added to the attributes of each corresponding equipment/pipeline. At that same time, the information of each sensor of the monitor system is input at the corresponding position, including fire information, environmental information, safety information and equipment information. And when a certain position is in Dan, that video information of the position can be automatically called. The above information is integrated through the BIM system to form an information-integrated BIM system. A manager can accurately position any structural part or equipment/pipeline through that visualization platform. Take a 360-degree view of what you want to know. For the equipment/pipeline in the power

system, in addition to obtaining the appearance and spatial location information, the platform covers different colors or patterns for the relevant fault parts to express their current different States (operation and maintenance state, cost control state, contract dispute state, etc.). Use this

intuitive way to assist managers in management and decision-making.

The application architecture of BIM-based power equipment operation and maintenance management visualization system is shown in Fig. 14.

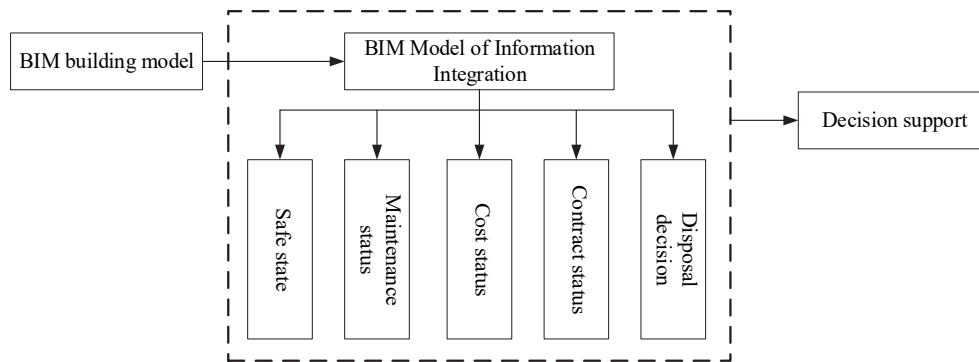


Figure 14 Application architecture of visualization system

7 CONCLUSIONS

As a crucial issue of electric power, equipment health management is studied with the perspective of big data in this paper. To achieve optimal control under the condition of lowest EFR and lowest energy consumption, the intelligent control system of power space environment based on EFR is designed from the perspective of multi-data fusion, EFR feature extraction and multi-objective optimal control. The research results help to realize multi-level centralized management, constant monitoring, intelligent warning and collaborative emergency response in urban power space. Although this paper integrates multi-source data from spatial management departments, there are still shortcomings in data dimensionality, and the main variables used for analysis are only six spatial environmental indicators. If the operational and environmental data of the data itself can be obtained, more effective and accurate characteristics of the EFR can be extracted. This paper relies mainly on single power spatial data. The accuracy of failure rate feature extraction cannot be guaranteed due to the short operation time and little equipment failure information, but the research system in this paper will be improved as time goes on.

Acknowledgements

This work is funded by Science and Technology Foundation of State Grid Xiongan Financial Technology Group co., Ltd (9100/2021-72001B).

8 REFERENCES

- [1] Mohan, T. R., Roselyn, J. P., Uthra, R. A., Devaraj, D., & Umachandran, K. (2021). Intelligent machine learning based total productive maintenance approach for achieving zero downtime in industrial machinery. *Computers & Industrial Engineering*, 157, 107267. <https://doi.org/10.1016/j.cie.2021.107267>
- [2] Ishii, H., Kawamura, K., Ono, T., Megumi, H., & Kikkawa, A. (1997). A fire detection system using optical fibres for utility tunnels. *Fire safety journal*, 29(2-3), 87-98. [https://doi.org/10.1016/S0379-7112\(96\)00065-3](https://doi.org/10.1016/S0379-7112(96)00065-3)
- [3] Namin, F. S., Ghafari, H., & Dianati, A. (2014). New model for environmental impact assessment of tunnelling projects. *Journal of Environmental Protection*, 2014. <https://doi.org/10.4236/jep.2014.56056>
- [4] Luo, Y., Alaghbandrad, A., Genger, T. K., & Hammad, A. (2020). History and recent development of multi-purpose utility tunnels. *Tunnelling and Underground Space Technology*, 103, 103511. <https://doi.org/10.1016/j.tust.2020.103511>
- [5] Rodríguez, J., Kennel, R. M., Espinoza, J. R., Trincado, M., Silva, C. A., & Rojas, C. A. (2011). High-performance control strategies for electrical drives: An experimental assessment. *IEEE transactions on industrial electronics*, 59(2), 812-820. <https://doi.org/10.1109/TIE.2011.2158778>
- [6] Canto-Perello, J. & Curiel-Esparza, J. O. R. G. E. (2006). An analysis of utility tunnel viability in urban areas. *Civil Engineering and Environmental Systems*, 23(1), 11-19. <https://doi.org/10.1080/10286600600562129>
- [7] Wang, P., Wang, S., Khan, M. I., & Zhu, C. (2020). Bending mechanics model and value of transverse joints in precast prestressed utility tunnel. *Journal of Asian Architecture and Building Engineering*, 19(3), 203-219. <https://doi.org/10.1080/13467581.2020.1743294>
- [8] Yoo, C., Jeon, Y. W., & Choi, B. S. (2006). IT-based tunnelling risk management system (IT-TURISK)—Development and implementation. *Tunnelling and Underground Space Technology*, 21(2), 190-202. <https://doi.org/10.1016/j.tust.2005.05.002>
- [9] Canto-Perello, J. & Curiel-Esparza, J. (2003, March). Risks and potential hazards in utility tunnels for urban areas. *In Proceedings of the Institution of Civil Engineers-Municipal Engineer*, 156(1), 51-56. Thomas Telford Ltd. <https://doi.org/10.1680/muen.2003.156.1.51>
- [10] Delanka-Pedige, H. M. K., Munasinghe-Arachchige, S. P., Abeywardana-Arachchige, I. S. A., & Nirmalakhandan, N. (2021). Wastewater infrastructure for sustainable cities: assessment based on UN sustainable development goals (SDGs). *International Journal of Sustainable Development & World Ecology*, 28(3), 203-209. <https://doi.org/10.1080/13504509.2020.1795006>
- [11] Wang, W., Zhu, Z., Jiao, Z., Mi, H., & Wang, Q. (2021). Characteristics of fire and smoke in the natural gas cabin of urban underground utility tunnels based on CFD simulations. *Tunnelling and Underground Space Technology*, 109, 103748. <https://doi.org/10.1016/j.tust.2020.103748>
- [12] Al-Kasasbeh, M., Abudayyeh, O., & Liu, H. (2021). An integrated decision support system for building asset

- management based on BIM and Work Breakdown Structure. *Journal of Building Engineering*, 34, 101959. <https://doi.org/10.1016/j.jobe.2020.101959>
- [13] Zhang, X. H., Guan, Y. X., Fang, Z., & Liao, Y. F. (2016). Fire risk analysis and prevention of urban comprehensive pipeline corridor. *Procedia Engineering*, 135, 463-468. <https://doi.org/10.1016/j.proeng.2016.01.156>
- [14] Chen, J., Jiang, L., Li, J., & Shi, X. (2012). Numerical simulation of shaking table test on utility tunnel under non-uniform earthquake excitation. *Tunnelling and Underground Space Technology*, 30, 205-216. <https://doi.org/10.1016/j.tust.2012.02.023>
- [15] Hingmire, S., Chougule, S., Palshikar, G. K., & Chakraborti, S. (2013, July). Document classification by topic labelling. *In Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*, 877-880. <https://doi.org/10.1145/2484028.2484140>
- [16] Maldonado-Bascón, S., Lafuente-Arroyo, S., Gil-Jimenez, P., Gómez-Moreno, H., & López-Ferrerías, F. (2007). Road-sign detection and recognition based on support vector machines. *IEEE transactions on intelligent transportation systems*, 8(2), 264-278. <https://doi.org/10.1109/TITS.2007.895311>
- [17] Yi, S., Huang, B., & Chan, W. T. (2005). XML application schema matching using similarity measure and relaxation labelling. *Information Sciences*, 169(1-2), 27-46. <https://doi.org/10.1016/j.ins.2004.02.013>
- [18] Sadollah, A., Eskandar, H., & Kim, J. H. (2015). Water cycle algorithm for solving constrained multi-objective optimization problems. *Applied Soft Computing*, 27, 279-298. <https://doi.org/10.1016/j.asoc.2014.10.042>
- [19] Kulkarni, M. N. K., Patekar, M. S., Bhoskar, M. T., Kulkarni, M. O., Kakandikar, G. M., & Nandedkar, V. M. (2015). Particle swarm optimization applications to mechanical engineering-a review. *Materials Today: Proceedings*, 2(4-5), 2631-2639. <https://doi.org/10.1016/j.matpr.2015.07.223>
- [20] Giagkiozis, I., Purshouse, R. C., & Fleming, P. J. (2015). An overview of population-based algorithms for multi-objective optimisation. *International Journal of Systems Science*, 46(9), 1572-1599. <https://doi.org/10.1080/00207721.2013.823526>
- [21] Norouzi, N., Tavakkoli-Moghaddam, R., Ghazanfari, M., Alinaghian, M., & Salamatbakhsh, A. (2012). A new multi-objective competitive open vehicle routing problem solved by particle swarm optimization. *Networks and Spatial Economics*, 12(4), 609-633. <https://doi.org/10.1007/s11067-011-9169-4>
- [22] Lalwani, S., Singhal, S., Kumar, R., & Gupta, N. (2013). A comprehensive survey: Applications of multi-objective particle swarm optimization (MOPSO) algorithm. *Transactions on combinatorics*, 2(1), 39-101.
- [23] Xue, G., Liu, S., Gong, D., & Ma, Y. (2019). Environmental Intelligent Control of Underground Integrated Pipe Gallery Based on Equipment Failure Rate. *IEEE Access*, 7, 160062-160071. <https://doi.org/10.1109/ACCESS.2019.2950381>

Contact information:**Xiaoming XIE**

(Corresponding author)

YGsoft Inc,

Guangwai Street, Xicheng District, Beijing, 100044, China

E-mail: xiexiaoming@ygsoft.com

Yixuan PENG

YGsoft Inc,

Guangwai Street, Xicheng District, Beijing, 100044, China

E-mail: pengyixuan@ygsoft.com