Hindawi
Science and Technology of Nuclear Installations
Volume 2022, Article ID 4172622, 14 pages
https://doi.org/10.1155/2022/4172622

Research Article

Study on Missing Data Filling Algorithm of Nuclear Power Plant Operation Parameters

Tianshu Wang, Ren Yu, and Qiao Peng

School of Nuclear Science and Technology, Naval University of Engineering, Wuhan 430033, China

Correspondence should be addressed to Ren Yu; 18071068480@163.com

Received 13 September 2021; Revised 8 December 2021; Accepted 13 January 2022; Published 4 February 2022

Academic Editor: Han Zhang

By analyzing the recorded operation data of a nuclear power plant (NPP), its results can serve the fault detection or operation experience feedback. Data missing exists in the recorded operation data. It may lower the data quality and affect the accuracy of the analysis results. In order to improve the data quality, two parts of researches are carried on. Firstly, to locate the missing data accurately the detecting algorithm for missing data of the NPP operation parameters based on wavelet analysis. Different judging basis is proposed for discrete and continuous missing respectively. Then, the filling method based on the hot deck algorithm are studied. As the dynamic properties of the parameters are closely related to the operating state of NPP, the similarity of the operation parameter vectors are formed to express the similarity of the operating states, so as to fulfill the requirements of the hot deck algorithm. To improve the accuracy of the measuring results, taken the differences between the characteristics of the analog parameters and the switch parameters into consideration, the similarity measurements using Mahalanobis distance for the analog parameter vectors and the matching measure for the switch parameter vectors are studied respectively. Finally, the operation data is taken to build the experiment data set for the algorithm verification. The results shows that the designed algorithm performs much better than the mean interpolation method and LSTM.

1. Introduction

With the development and application of digital instrument and control (I&C) system in nuclear power plant (NPP), the capacity of data storage and analysis is improved. Many big data analysis algorithms, like Machine Learning and Deep Learning, can be utilized to analyze the operation data [1]. Most of the operation data collection, transmission and storage in NPP are carried out by automatic instruments. Due to the influence of environment interference, inherent characteristics of instruments and some other reasons, the NPP operation data missing may occur randomly, which may directly affect the analysis of data. Therefore, the research on methods to improve data quality, which are called data cleaning, has received widespread attention [2]. Data cleaning aims to identify and correct the noise in the data and minimize its impact on the data analysis results. Therefore, before the application of big data analysis algorithm in NPP, it is necessary to study the detecting and filling algorithm of missing data in operation parameters. Noise in data mainly includes data missing, redundant data, conflicting data, and wrong data, which are collectively called dirty data [3, 4]. Currently, the widely studied data cleaning methods include methods based on removal, direct manipulation, models, and imputation [5].

The methods based on removal will reduce the amount of records used in the training process or in the adjustments of the prediction model when mining the data of NPP. As a result, the ability of finding consistent patterns gets weak [6]. The methods based on direct manipulation are commonly nonparametric. Therefore, when applied to NPP operating data, it makes the correlation of the attributes lower, which negatively influence the performance of the algorithms [7]. The methods based on models include statistical, probabilities and learning techniques for obtaining a model. The iterative algorithm can take quite a long time before it converges with the growth of the amount of data [8]. As a result, if the NPP operating data were processed with this

algorithm, the large amount of data would lead to a low efficient. Compared with the above algorithms, the method based on imputation has the advantages of lower computational complexity, higher computational efficiency and higher accuracy. These advantages make it have advantages in dealing with the large amount of data and high correlation of the NPP operating data.

Scholars have conducted a large number of relevant studies about filling algorithm for missing data based on imputation. Ragel and Crémilleux proposed an imputation method using the RAR-Robust Association Rules algorithm in [9]. Then, in [10], the application of this method in a real database is demonstrated. In [11], a method based on Formal Concept Analysis is brought out. It uses an implication basis, which represents the dependencies between attributes to impute value for missing data. In [12], a composite imputation is proposed which applied other tasks, including the data clustering and the attribute selection, before the imputation process of a missing data to improve the quality of the imputed data. In [13], the Radial Basis Function Network classifier is applied to improve the data quality in data sets containing missing data. Wu et al. found a practical utilization of association rules to complete missing data in [14]. In [15], in order to extract the characteristics of motor signals affected by noise, a new advancing coupled multistable stochastic resonance method, namely CMSR, is proposed to deal with the dirty data.

However, most studies on the imputation methods tend to be more theoretical. L. O. Silva points out in [5] that in the literature of data mining context, few studies take the data missing into consideration. When it comes to data mining in NPP, the problem of data missing also remained to be solved and only a few researches has been done. The most representative research is conducted in [16]. A missing data imputation algorithm based on least squares support vector machine (LSSVM) is proposed to reconstruct the missing data in environmental radiation monitor sensor network of NPP. This research has very high reference value. It lays a good foundation for the subsequent research on data missing of NPP. The K-means clustering algorithm based on a noise algorithm in [17] provides a new way of thinking and comparison. And the mode of data analysis in [18] gives good references to time series data analysis in NPP.

Therefore, to deal with the operation data missing in NPP, the detecting algorithm based on wavelet analysis, and the filling method based on hot deck algorithm and Mahalanobis distance are studied, referring to the research achievements in other fields. In Section 1, the characteristics of the NPP operation data are analyzed, and the filling strategy based on hot deck algorithm is brought up. In Section 2, the missing data detecting algorithm based on wavelet decomposition is designed, which provides a solid foundation for the identification of missing data. In Section 3, the construction method for NPP operation state vector and the corresponding similarity measurement based on Mahalanobis distance and matching measure are designed. The detailed algorithm for filling missing data of NPP is further designed in Section 4. Taking the operation data of a NPP as a sample, the application effect of the designed algorithm is verified in Section 5. And the result is analyzed in Section 6.

2. The Characteristics of NPP Operation Data and the Data Filling Strategy

2.1. Characteristics Analysis of NPP Operation Parameters. The operation parameters of NPP are commonly divided into two types, the analog parameter and the switch parameter. And these parameters are coupled together by complex physical relationships. It indicates that the parameters are not independent from each other. The operation state of the NPP at a certain time can be represented by the two types of parameters. In another word, a particular operation state of a NPP must correspond strictly to a set of operation parameter value. These operation parameters representing the operation state of NPP are constructed into a vector, which is called the operation state vector of a NPP, or state vector for short. Due to the influence of various random errors in data collecting and transmitting, the value of the parameters may not be exactly the same. However, the similarity is still extremely high. Therefore, the missing data of NPP can be filled by searching the similar system operation state.

Data missing of NPP refers to the abnormal returning to zero of the operation data collected by the instrument and control (I&C) system, due to the interference of the external environment or the random fault of the data acquisition and recode device. Data missing is represented in two forms.

(1) Discrete missing: it refers to the abnormal return to zero of discrete points in the process of a parameter changing with time.

(2) Continuous missing: it refers to the abnormal return to zero of a number of continuous points in the process of a parameter changing with time.

There are differences between the mathematical characteristics of the two kinds of data missing. If there is a discrete missing in a parameter data recode, the missing point is represented as the first kind of discontinuity point in the time function of the parameter. If there is continuous missing in a parameter data recode, the first and last points of the missing segment are represented as the first kind of discontinuous points in the time function of the parameter.

2.2. The Missing Data Filling Algorithm. The requirements of the filling algorithm for data missing are further clarified as follows:

(1) The algorithm should have the ability to reduce using algorithms that require verification of parameter data independence.

(2) The designed algorithm should make full use of the corresponding relationship between the operation data of NPP and its operation state.

According to the requirements above, the alternative algorithms include the hot deck (HD) algorithm, the KNN algorithm, and the regression replacement algorithm [19].

For an event with missing data, the HD algorithm tries to find another event with complete data, which is the most similar to it. And then, the missing data of the previous event is filled with the corresponding data of the found event. Different events may use different similarity measurements [20].

KNN is the most typical representative of clustering methods. Firstly, the nearest k samples of missing data are determined according to the distance (commonly used Euclidean distance), and the weighted average of these k values is used to estimate the missing data of this sample.

The regression filling algorithm needs to select a number of independent parameters first. And then establish a regression equation with the parameters to estimate the missing value. That is to say, the missing value is replaced by the conditional expected value of the missing data [21]. The artificial neural network method is the most representative one of it [22].

KNN is a special case of the hot deck algorithm. The weighted average of k groups of data will passivate the influence of non-measured values associated with the parameters, which is not conducive to the filling effect. The filling effect of regression filling algorithm depends on the accuracy of the regression equation. Due to the fact that the operation data of NPP is complex and changeable, the establishment of regression equation has high requirements on data quantity and computer calculation force, which is not suitable for the current application. Magnani points out in [23], some advantages of the hot deck: reduction of the standard error without imposing a rigid model; production of a data set without missing data, and preservation of population distribution. Moreover, this method allows distinct imputation techniques to be used for each group generated.

The HD algorithm is simple in concept and uses the relationship between operation parameters and the operation state of NNP to estimate the missing data. Therefore, the HD algorithm is selected to deal with the missing data of NPP.

The HD algorithm for the missing data of the operation parameters of NPP mainly consists of the following steps:

(1) Detection of missing points in the operation data.

(2) Generation of the operation state vector.

(3) Calculation of the operation state similarity.

(4) Select the time point with perfect data which is most similar to the time with missing.

(5) Fill the missing data with the value of the corresponding parameter in the time point with perfect data.

(6) Check the switch parameter affected by the missing data and correct the wrong ones.

The brief flow of the filling algorithm is shown in Figure 1.

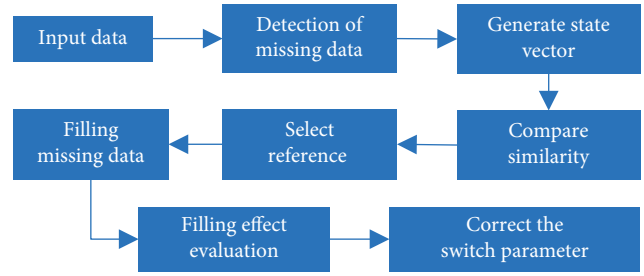

Figure 1: Brief flow chart of the filling algorithm.

3. Design of Missing Data Detection Algorithm

The data missing of NPP is usually manifested as a numerical mutation of 0, but the case where the data naturally turns to 0 should be ruled out. That is, a data value of 0 is a necessary and insufficient condition for the point to be a data missing point. Therefore, it is necessary to further design the algorithm to detect whether the data return to 0 is a data missing point.

The analog parameters of NPP are continuous and derivable in the time domain. According to the analysis of data characteristics, the premise of determining whether a zero is a missing point is to determine whether its characteristics meet the requirements of discontinuity points. Here, the wavelet decomposition algorithm [24] is adopted to detect the discontinuity point.

Assume that the scale function and wavelet function of a certain wavelet are Φ and Ψ. For any non-negative integer j, V_j is defined as a j-order step function space spanned by the following set of functions as formula (1) over the real number field:

$$\left\{\ldots, \Phi\left(2_{t+1}^{j}\right), \Phi\left(2_{t}^{j}\right), \Phi\left(2_{t-1}^{j}\right), \left(2_{t-2}^{j}\right), \ldots\right\}, \tag{1}$$

where, t stands for the time label.

Define:

$$w_j = \sum_{k\in Z} a_k \Psi\left(2^j t - k\right), \tag{2}$$

where, a_k is the coefficient of the function $a_k \in R$.

According to the wavelet decomposition theory, for any function f with finite discontinuous points, a step function $f_j \in V_j$ can be used to approximate it infinitely. If j is large enough, $f_j(t)\, f(t)\, f_j$ can then be further decomposed into the sum of its sub-projection spaces, that is:

$$f_j = f_0 + w_{j-1} + w_{j-2} + \cdots + w_0. \tag{3}$$

Assume that the function of the analog parameter x and time t is $x = f(t)$. To detect the discontinuous points in $f(t)$ with wavelet, the above algorithm is used. It means 2^j segmentation points are inserted into the detection interval uniformly spaced to discretize it. In this case, $f_j(t)$ can be represented as follows:

$$f_j(t) = \sum_{k=0}^{2^j} a_k^{(j)} \Phi\left(2^j t - k\right), \tag{4}$$

where $a_k^{(j)} = f\left(k/2^j\right)$ and $f_j(t) \in V_j$. It is further decomposed into the sum of $f_{j-1}(t)$ and $w_{j-1}(t)$, where $f_{j-1}(t) \in V_{j-1}$ and $w_{j-1}(t) \in w_{j-1}$.

The projection coefficient of $f_j(t)$ on the lower order wavelet space w_{j-1}, namely, the form of the wavelet

coefficient $a_k^{(j-1)}$ is half of the difference between the adjacent coefficients of the higher order step function, i.e. $1/2\,(a_{2k}^{(j)} - a_{2k+1}^{(j)})$. Thus, if there is a discontinuous point $f(t_0)$ at $(t = t_0\,(k/2^j) \le t_0 < k + 1/2^j)$, the absolute value of $b_k^{(j-1)}$ will be relatively large due to the great difference between the values of $f\,(k/2^j)$ and $f\,(k + 1/2^j)$ on both sides of the discontinuity point t_0. The difference between $f\,(k/2^j)$ and $f\,(k + 1/2^j)$ at the continuous point is very small, so $a_k^{(j-1)}$ will also be very small, close to zero. Thus, whether $f(t)$ is continuous at t_0 can be judged based on the value of $a_k^{(j-1)}$. The analog parameter curve of the NPP is decomposed by wavelet, and the set $a_k^{(j-1)}$ is reconstructed. If there is a discontinuity point of the first kind in the original waveform, the corresponding point of the reconstructed waveform will be a distortion peak.

In essence, the above derivation process aims to extract the first derivative characteristics of the data by wavelet analysis, so as to judge its continuity. Therefore, it is not necessary to specify the selected wavelet basis, and any common wavelet basis function can meet the requirements of the design algorithm in this paper. At the same time, the first-order derivative of the function can be satisfied only by first-order decomposition.

As shown in Figure 2, a discrete missing appears at the time point 4000 of a certain parameter, and an obvious characteristic peak appears at the corresponding time point in Figure 3 of its reconstruction model.

As shown in Figure 4, (a) continuous data missing appears at time 4000 to 5000. In its reconstruction model of Figure 5, the corresponding beginning and end, time 4000 and 5000, show obvious characteristic peaks.

To sum up, the detection algorithm for missing data of NPP is listed as follows:

(1) Scan the data values of operation parameters and mark all zeros.

(2) Aggregate the continuous zeros into segments.

(3) Decompose the data by wavelet and reconstruct its high frequency terms to obtain the derivative characteristic waveform.

(4) Check the discrete and continuous zeros:

(A) If it is a discrete zero, check whether there is a distortion peak at the time point corresponding to its characteristic waveform; if so, this point is a discrete missing point.

(B) If it is a continuous zero, check whether there is a distortion peak at the beginning and end of the characteristic waveform of the corresponding segment. If the two characteristic peaks appear, at the beginning and end of the segment, it determines that the segment is continuous missing, and all points in the segment are missing points.

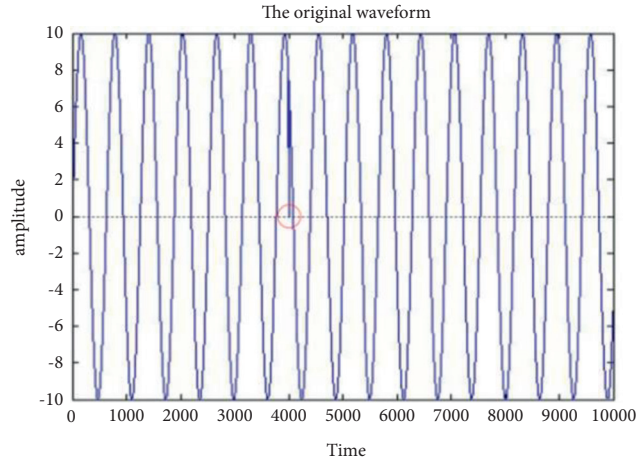

Figure 2: Waveform of the sample data.

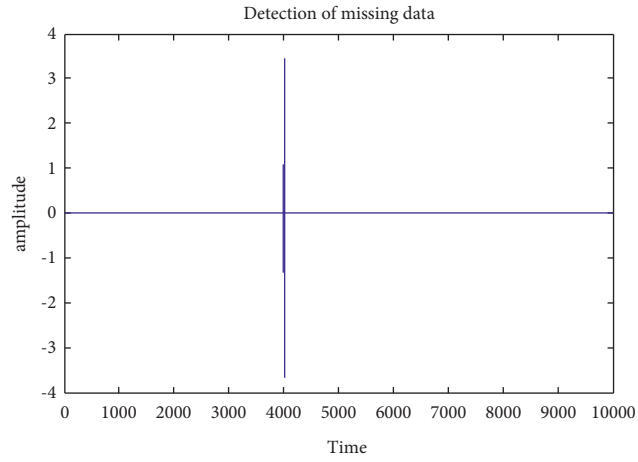

Figure 3: Detection waveform of the sample data.

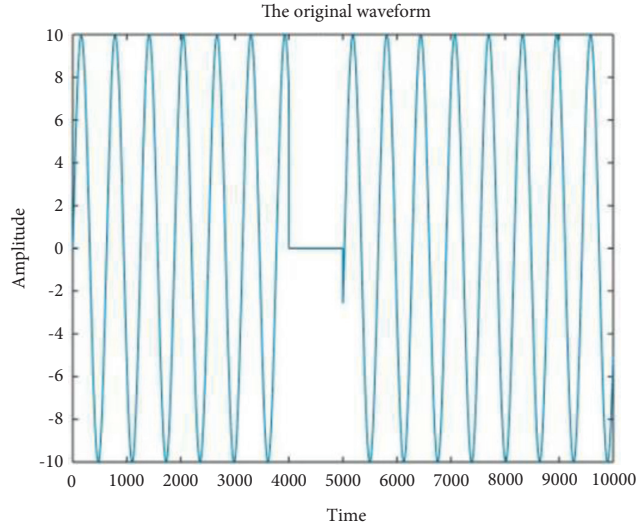

Figure 4: Waveform of the sample data.

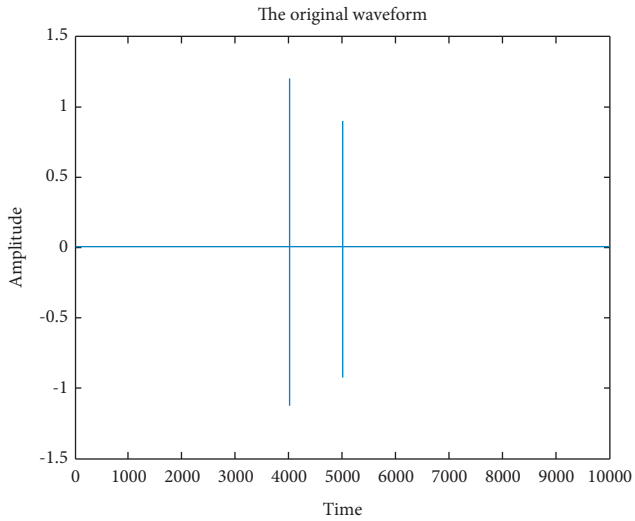

Figure 5: Detection waveform of the sample data.

4. Design of Operation State Vector and the Similarity Measurement

4.1. Design of Operation State Vector. To deal with the missing operation data of NPP using HD algorithm, the system operation state vector should be constructed first. Therefore, the data similarity measurements of NPP operation data are studied.

There are obvious differences between the operation parameters of the analog and switch parameters of the NPP, if the two kinds of parameters are grouped into the same state vector, the similarity of the vector measured by a single method will not only increase the computational complexity of the algorithm.

To completely represent the operation state of the NPP, all parameters of NPP are used to form vectors to represent the operation state of the NPP. In order to accurately and completely retain the operating state information of the NPP, the assignment of elements in the designed operating state vector are absolutely original collected data without any process.

4.1.1. Switch Parameter State Vector $X_b(t)$. Switch parameter state vector, $X_b(t)$, is a time-varying function vector, where t stands for the time point. It is composed of all switch parameters, and each element x_{b_i} in the vector represents a switch or alarm parameter. The element value $x_{b_i}(t)$ represents the measured value of the corresponding switch parameter at time t. $X_b(t)$ is shaped as below:

$$X_b(t) = \left\{x_{b_1}(t), x_{b_2}(t), x_{b_3}(t), \ldots, x_{b_l}(t)\right\}, \tag{5}$$

where $x_{b_i}(t) = 0$ or 1, $(1 \le i \le l)$, and l is equal to the sum of the number of switch parameters and the number of alarm parameters. And t stands for the time label.

4.1.2. Analog Parameter State Vector $X_r(t)$. Analog parameter state vector, $X_r(t)$, is a time-varying function vector, where t stands for the time point. It is composed of all analog parameters, and each element x_{r_j} in the vector represents an analog parameter. The element value $x_{r_j}(t)$ represents the measured value of the corresponding switch parameter at time t. $X_r(t)$ is shaped as below:

$$X_r(t) = \left\{x_{r_1}(t), x_{r_2}(t), x_{r_3}(t), \ldots x_{r_s}(t)\right\}, \tag{6}$$

where $x_{r_j}(t) \in R$, $(1 \le j \le s)$, and s is equal to the sum of the number of analog parameters. And t stands for the time label.

4.2. Design of the Similarity Measurement for Switch Parameters State Vector. A feature is called binary when there are only two states (0, 1) for it, and 0 means false, while 1 means true. All Boolean-type parameters are of typical binary feature. The match degree is used here to characterize the similarity of the Boolean parameter vector $X_b(t)$.

4.2.1. Definition of Match. For two components, x_i and y_i, of a given vector, $\vec{x}$ and $\vec{y}$,

(1) if $x_i = 1$ and $y_i = 1$, then it is called a 1–1 match;

(2) if $x_i = 0$ and $y_i = 1$, then it is called a 0–1 match;

(3) if $x_i = 1$ and $y_i = 0$, then it is called a 1–0 match;

(4) if $x_i = 0$ and $y_i = 0$, then it is called a 0–0 match.

4.2.2. Selection of Matching Measure. In common methods of matching measure, the simple matching measure can perfectly fit the requirement of the similarity measurement of Boolean-type parameter state vector. Therefore, the simple matching measure is used to measure the similarity of Boolean-type parameter state vectors of NPP.

Assume that there are two Boolean parameter vectors X_{b_i}, X_{b_j} $(1 \le i, j \le l)$. Let:

$$a = \sum_k X_{b_{i_k}} X_{b_{j_k}}, \tag{7}$$

which represents the number of 1–1 matches between X_{b_i} and X_{b_j}, and:

$$e = \sum_k \left(1 - X_{b_{i_k}}\right)\left(1 - X_{b_{j_k}}\right), \tag{8}$$

which represents the number of 0–0 matches between X_{b_i} and X_{b_j}. Then the simple match degree is calculated as formula (9):

$$m(\vec{x}, \vec{y}) = \frac{a + e}{n}. \tag{9}$$

According to its definition, m $(0 \le m \le 1)$ is a real number, and the greater the value of m is, the higher the similarity is.

4.3. Design of the Data Similarity Measurement for Analog Parameters State Vector. The similarity measurements of analog parameter state vectors in pattern recognition are

distance measure and angle measure. The distance measure uses generalized distance to represent the similarity between vectors. The smaller the distance is, the higher the similarity is. The angle cosines are commonly used to represent the similarity between vectors. According to the monotonicity of the cosine function, the higher the cosine value is, the higher the similarity is.

For the analog parameter state vectors of NPP, if the angle cosine is used to represent the similarity between them, the length information will be lost, resulting in data distortion. For example, the cosine value of vectors, $\vec{x}$ = (2, 2, 4) and $\vec{y}$ = (1, 1, 2), is 1, which means they are of very high similarity. However, if such situation occurs in the operation of NPP, it is obvious that they are in two different operation states with low similarity. Therefore, the distance measure is chosen as the analog parameter state vector similarity measurement.

At present, the commonly used distance measures include Euclidean distance, Mahalanobis distance, Minkowski distance, Chebyshev distance, etc. The dimension of different operation parameters of NPP is different. Therefore, the distance measuring method is required to weaken the influence of different dimensions. Mahalanobis distance can take the relationship between parameters into account, and it is independent of dimension.

Let the data set DT be a vector set composed of X_r corresponding to T continuous time points in the NPP. $DT(s) = X_r(t_s)$, $1 < s \leq T$. x and y are two vectors of different time points in DT, The calculation formula of Mahalanobis distance between x and y is shown as formula (10):

$$D_M = \sqrt{(x-y)^T \sum^{-1} (x-y)}, \tag{10}$$

where $\sum$ represents the covariance matrix of DT, and $\sum^{-1}$ represents the inverse matrix of $\sum$, where $\sum$ is calculated through formula (11).

$$\sum = \begin{pmatrix} c_{11} & \cdots & c_{1s} \\ \vdots & \ddots & \vdots \\ c_{s1} & \cdots & c_{ss} \end{pmatrix},$$
$$c_{ij} = \mathrm{Cov}\left(X_{r_i}, X_{r_j}\right), \quad X_{r_i}, X_{r_j} \in DT, \tag{11}$$

where $f(x, y) = \mathrm{Cov}(x, y)$ stands for the covariance between vectors x and y.

4.4. Design of Synthetic Similarity Calculation Algorithm. The similarity measurement for the operation parameters of NPPs with different data types is proposed. The two methods are synthesized into a unified similarity measurement for system operation state, which provides the calculation basis of HD algorithm.

There are three types of operation parameters in a NPP, named analog parameters, switch parameters and alarm parameters. The relation.

(1) The states of the alarm parameters are based on the judgment of analog parameters' threshold value, so the value of the analog parameters reflects the states of the alarm parameters to a large extent.

(2) The switch parameters indicate the operation states of the equipment, such as pumps and valves in a NPP, and the changes of the equipment state will cause the change of the analog parameters.

It can be seen that the analog parameters contain some information about alarm parameters and switch parameters. Therefore, the similarity measurement of the NPP is designed at different levels. In detail:

(1) First-level description. The Mahalanobis distance is calculated by using analog parameters. The data set with the smallest Mahalanobis distance is the most similar one of the data to be filled.

(2) Second-level description. If multiple groups of the same minimum Mahalanobis distance appear in the calculation results of the first level description, the simple matching measure of the Boolean-type parameter state vector is calculated and arranged in ascending order according to the calculation results, the largest group of data is the data group most similar to the data to be filled.

5. Design of Filling Algorithm for Missing Data

5.1. Design of Algorithmic Flow. The specific process of constructing the missing data filling algorithm for the operation parameter data of NPP is as follows:

5.2. Analysis of the Computation Complexity for the Algorithm. Suppose there are M parameters in the data set, and each parameter has N data points, in which there are K missing points in total. The missing points in the data set are much smaller than the size of the data set. At the same time, the number of the operating parameters is also much smaller than the size of the data set. Therefore, it can be concluded as formula.

$$M, K \ll N. \tag{12}$$

Both M and K can be regarded as constants. The computational complexity of the entire algorithm is shown in equation (13).

Final computational complexity
= MAX{computational complexity of missingvalue detection, computational complexity of missing value filling}. (13)

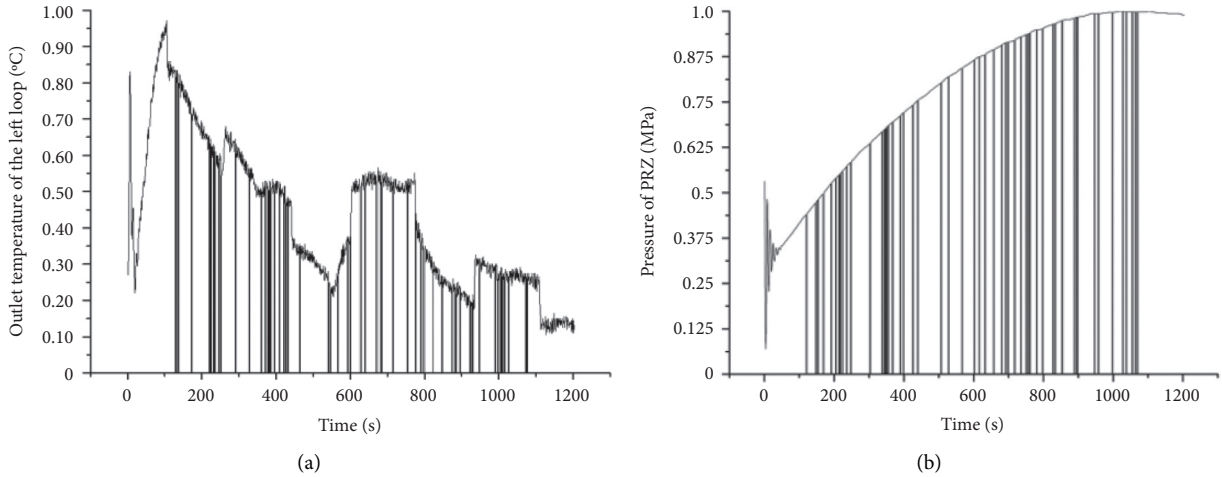

Figure 6: The curves of parameters after setting missing data.

The computational complexity of missing value detection mainly depends on the computational complexity of the calculation process for wavelet decomposition. In general, the computational time complexity of the wavelet decomposition algorithm is O(Nlog(N)). Therefore, the computational complexity of missing value detection is K·O(Nlog(N)), that is, O(Nlog(N)).

Computational complexity of missing value filling is shown in equation (14)

$$\text{Computational complexity of missing value} = K \cdot M \cdot N \cdot \text{Tmp}, \tag{14}$$

where, Tmp represents the computational complexity required to calculate the Mahalanobis distance, and formula (15) calculates Tmp.

$$\text{Tmp} = M^2 O(N). \tag{15}$$

In summary, the computational complexity of missing value filling is O(N^2). Therefore, the final computational complexity of the algorithm designed in this paper is O(N^2). Compared with common algorithms, the designed algorithm is of high efficiency.

6. Experimental Verifications of the Proposed Algorithm

To verify the correctness and advantages of the designed algorithm, the operation data from the simulator of CAP 1400 developed by SJTU after normalization is used as a sample to carry out the experiments.

In order to make the experimental results easier to be analyzed, all the data are shown in a normalized form. A data set of 1200 time points after setting data missing is taken as the sample. The Error Rate is used to evaluate the effect of the algorithms. The calculation of the error rate is shown as formula (16):

$$E = \frac{|F - T|}{T} \times 100\%, \tag{16}$$

where, E is the Error Rate, F is the result calculated by the algorithms, and T is the true value which the missing points actually should be.

The filling methods based on mean interpolation (MI) and the method based on LSTM are set as the compare algorithms, to verify the advantage of the designed algorithm. For the MI algorithm, data near the missing point are selected as the calculation basis, and the average value is used as the filling value of the missing point. For the LSTM algorithm, several data before the missing point are selected as the basis, use the LSTM method for prediction, and then the predicted value is used as the filling value. Calculate the error rate of the three algorithms for the same missing point by formula (16) and compare the calculation results to reflect the superiority of the designed algorithm.

6.1. Experiment on Data with Discrete Missing Points. 60 discrete missing points are set into two parameters, the outlet temperature of the first loop and the pressure of the Pressurizer (PZR), randomly. The curves of the parameters after setting are shown in Figure 6.

The detection results of the missing data are shown in Figure 7.

The curves of data after filling using the designed algorithm are shown in Figure 8.

Taking the filling data of the first loop outlet temperature as an example, the error rates of a part of calculation results are listed and compared with the other two algorithms in Table 1. And the full vision of the data is listed in the Attached Table 2 and Attached Table 3.

6.2. Experiment on Data with Continuous Missing Points. Three sections of continuous missing are set into the pressure of 1# steam generator (SG), and the duration of the three sections are set as 10 seconds, 50 seconds and 100 seconds. The curve after setting missing is shown in Figure 9. Missing data is detected, and the results are shown in Figure 10.

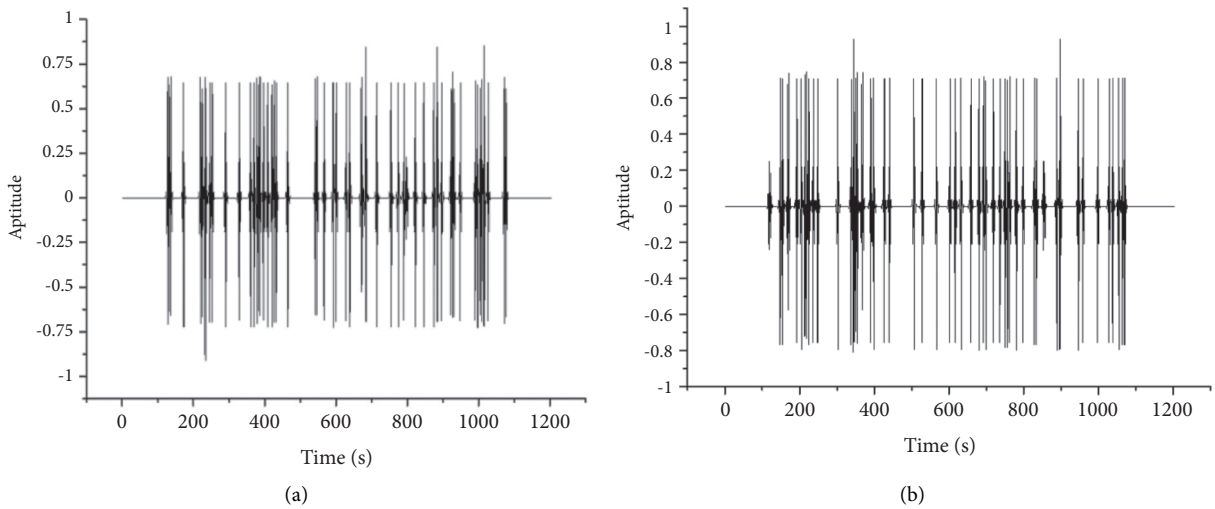

Figure 7: Detection results of the missing data.

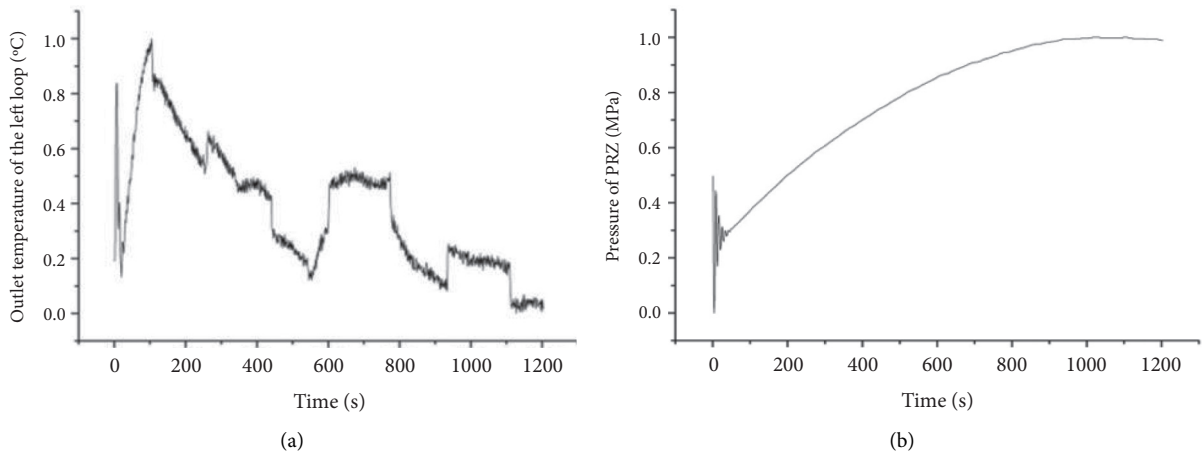

Figure 8: The curves of the parameters after filling.

(i) **Input:** The operation data set
(ii) **Output:** The replacement data
(iii) Sum = The total number of parameters;
(iv) T = The time length of dataset;
(v) for $i = 1$ to Sum do
(vi) if (there is value 0 in Parameter[i]) then
(vii) Integrate the continuous zero points of each parameter into zero intervals of Parameter[i];
(viii) Decompose the time curve of Parameter[i] by wavelet;
(ix) Reconstruct the high frequency terms;
(x) Detect discrete missing and continuous missing;
(xi) Mark the missing points;
(xii) Num = The total number of parameters with data missing points;
(xiii) Establish the switch parameters state vector;
(xiv) Establish the analog parameters state vector;
(xv) for $i = 1$ to Num do
(xvi) Select the i^{th} parameter with missing data;
(xvii) Tmp = The total number for the missing data points of the selected data;
(xviii) for $j = 1$ to Tmp do
(xix) Select the j^{th} missing point as the time point to be filled;

Algorithm 1: Continued.

(xx) For $k = 1$ to T do
Calculate the Mahalanobis distance between the point to be filled and the complete ones;
If current Mahalanobis distance < Minimum
Refresh Minimum;
Record the corresponding data;
else if Mahalanobis distance = = Minimum
Calculate the matching measure between the switch parameter state vectors;
Select the more similar one;
Refresh Minimum;
Record the corresponding data;
Fill the missing points with the recorded data;
Check the related switch parameter and correct the wrong ones;
Return Dataset after filling

Algorithm 1: Filling missing data.

Table 1: Comparison of the filling error rates for missing data in outlet temperature of the first loop.

No.	MI	LSTM	HD	No.	MI	LSTM	HD	No.	MI	LSTM	HD
10	7.98E-06	4.47E-05	3.54E-06	20	2.01E-16	4.47E-05	3.54E-06	51	1.77E-06	5.6E-05	1.77E-06
11	1.77E-06	4.47E-05	5.32E-06	26	7.09E-06	4.47E-05	3.54E-06	52	7.09E-06	4.47E-05	0
12	1.77E-06	4.47E-05	0	27	7.09E-06	4.47E-05	5.32E-06	53	1.77E-06	1.14E-05	1.77E-06
13	7.98E-06	7.81E-05	7.09E-06	28	8.86E-07	4.47E-05	7.09E-06	54	1.77E-06	5.46E-05	1.77E-06
14	2.66E-06	3.48E-05	5.32E-06	29	1.77E-06	4.47E-05	0	55	2.66E-06	2.17E-05	1.77E-06
15	1.77E-06	3.44E-05	0	30	3.54E-06	4.47E-05	1.77E-06	56	3.54E-06	3.26E-05	1.77E-06
16	1.77E-06	3.34E-05	3.54E-06	47	3.54E-06	4.47E-05	1.77E-06	57	3.54E-06	3.16E-05	3.54E-06
17	2.66E-06	4.47E-05	1.77E-06	48	3.54E-06	1.14E-05	1.77E-06	58	1.15E-05	7.62E-05	5.32E-06
18	3.54E-06	4.47E-05	0	49	5.32E-06	5.46E-05	1.77E-06	59	8.86E-07	6.63E-05	0
19	1.77E-06	4.47E-05	7.09E-06	50	8.86E-07	5.5E-05	1.77E-06	60	7.09E-06	2.27E-05	3.54E-06

Table 2: (Attached table) Comparison of the filling error rates for missing data in outlet temperature of the first loop.

No.	MI	LSTM	HD	No.	MI	LSTM	HD	No.	MI	LSTM	HD
1	2.11E-05	4.47E-05	7.09E-06	21	8.03E-06	4.47E-05	2.01E-16	41	5.55E-06	1.139E-05	2.01E-16
2	5.28E-06	4.47E-05	2.66E-06	22	5.36E-06	4.47E-05	7.98E-06	42	3.34E-05	2.12E-05	3.54E-06
3	1.58E-05	4.47E-05	1.77E-06	23	5.36E-06	4.47E-05	2.66E-06	43	1.55E-16	6.492E-05	4.43E-06
4	1.32E-05	4.47E-05	8.86E-06	24	1.07E-05	4.47E-05	1.77E-06	44	5.57E-06	6.63E-05	5.32E-06
5	1.05E-05	4.47E-05	5.32E-06	25	0	4.47E-05	0	45	2.79E-06	5.601E-05	5.32E-06
6	2.65E-06	4.47E-05	2.66E-06	26	7.09E-06	4.47E-05	3.54E-06	46	2.79E-06	4.472E-05	7.09E-06
7	1.06E-05	4.47E-05	6.20E-06	27	7.09E-06	4.47E-05	5.32E-06	47	3.54E-06	4.47E-05	1.77E-06
8	0	4.47E-05	0	28	8.86E-07	4.47E-05	7.09E-06	48	3.54E-06	1.14E-05	1.77E-06
9	2.11E-05	4.47E-05	7.09E-06	29	1.77E-06	4.47E-05	0	49	5.32E-06	5.46E-05	1.77E-06
10	7.98E-06	4.47E-05	3.54E-06	30	3.54E-06	4.47E-05	1.77E-06	50	8.86E-07	5.5E-05	1.77E-06
11	1.77E-06	4.47E-05	5.32E-06	31	2.72E-06	4.47E-05	3.54E-06	51	1.77E-06	5.6E-05	1.77E-06
12	1.77E-06	4.47E-05	0	32	0	4.47E-05	2.66E-06	52	7.09E-06	4.47E-05	0
13	7.98E-06	7.81E-05	7.09E-06	33	1.51E-16	4.47E-05	8.86E-07	53	1.77E-06	1.14E-05	1.77E-06
14	2.66E-06	3.48E-05	5.32E-06	34	2.73E-06	4.47E-05	5.32E-06	54	1.77E-06	5.46E-05	1.77E-06
15	1.77E-06	3.44E-05	0	35	2.74E-06	4.47E-05	3.54E-06	55	2.66E-06	2.17E-05	1.77E-06
16	1.77E-06	3.34E-05	3.54E-06	36	2.74E-06	1.139E-05	9.75E-06	56	3.54E-06	3.26E-05	1.77E-06
17	2.66E-06	4.47E-05	1.77E-06	37	2.74E-06	5.462E-05	2.01E-16	57	3.54E-06	3.16E-05	3.54E-06
18	3.54E-06	4.47E-05	0	38	2.48E-05	5.503E-05	2.66E-06	58	1.15E-05	7.62E-05	5.32E-06
19	1.77E-06	4.47E-05	7.09E-06	39	5.53E-06	5.601E-05	2.66E-06	59	8.86E-07	6.63E-05	0
20	2.01E-16	4.47E-05	3.54E-06	40	8.31E-06	4.472E-05	2.66E-06	60	7.09E-06	2.27E-05	3.54E-06

Table 3: (Attached table) Comparison of the filling error rates for missing data in pressure of PRZ.

No.	MI	LSTM	HD	No.	MI	LSTM	HD	No.	MI	LSTM	HD
1	7.09E-06	1.998E-04	6.45E-06	21	2.01E-16	2.334E-04	6.45E-06	41	0.500002	1.763E-04	9.02E-05
2	2.66E-06	2.413E-04	0	22	7.98E-06	1.671E-04	0	42	3.54E-06	2.180E-04	0
3	1.77E-06	2.417E-04	1.29E-05	23	2.66E-06	2.169E-04	1.93E-05	43	4.43E-06	2.184E-04	0
4	8.86E-06	2.083E-04	0	24	1.77E-06	2.507E-04	1.93E-05	44	5.32E-06	2.183E-04	6.44E-06
5	5.32E-06	2.080E-04	5.16E-05	25	0	2.509E-04	6.44E-06	45	5.32E-06	1.763E-04	0
6	2.66E-06	2.166E-04	5.16E-05	26	5.32E-06	2.003E-04	6.44E-06	46	7.09E-06	1.847E-04	0
7	6.20E-06	2.251E-04	6.45E-06	27	7.09E-06	2.000E-04	6.44E-06	47	7.09E-06	1.602E-04	0
8	0	2.251E-04	6.45E-06	28	7.09E-06	1.754E-04	0	48	3.54E-06	2.106E-04	0
9	1.77E-06	2.251E-04	6.45E-05	29	8.86E-07	1.923E-04	0	49	3.54E-06	2.444E-04	6.44E-06
10	7.98E-06	2.251E-04	0	30	1.77E-06	2.010E-04	1.29E-05	50	5.32E-06	2.359E-04	0
11	1.77E-06	2.251E-04	6.45E-06	31	3.54E-06	2.097E-04	0	51	8.86E-07	1.850E-04	0
12	1.77E-06	2.251E-04	1.29E-05	32	0.500001	2.097E-04	6.44E-06	52	1.77E-06	1.847E-04	0
13	7.98E-06	1.918E-04	1.93E-05	33	0.500003	2.097E-04	6.44E-06	53	7.09E-06	1.602E-04	0
14	2.66E-06	2.334E-04	6.45E-06	34	5.32E-06	1.763E-04	6.44E-06	54	1.77E-06	2.106E-04	0
15	1.77E-06	2.338E-04	4.51E-05	35	3.54E-06	2.180E-04	0	55	1.77E-06	1.777E-04	0
16	1.77E-06	2.337E-04	6.45E-06	36	9.75E-06	2.184E-04	7.08E-05	56	2.66E-06	1.861E-04	0
17	2.66E-06	2.251E-04	1.93E-05	37	2.01E-16	2.183E-04	1.29E-05	57	3.54E-06	1.861E-04	0
18	3.54E-06	1.918E-04	1.29E-05	38	2.66E-06	2.097E-04	0	58	3.54E-06	1.949E-04	0
19	1	2.334E-04	1.93E-05	39	2.66E-06	2.097E-04	6.44E-06	59	1.15E-05	1.949E-04	0
20	7.70E-05	2.338E-04	1.29E-05	40	0.500001	2.097E-04	6.44E-06	60	8.86E-07	1.949E-04	0

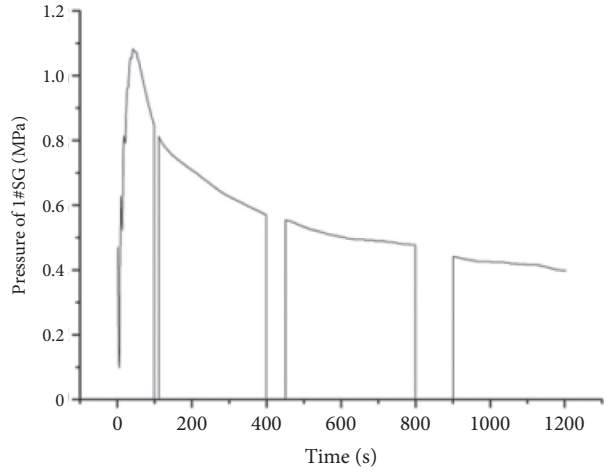

Figure 9: Curve after setting missing.

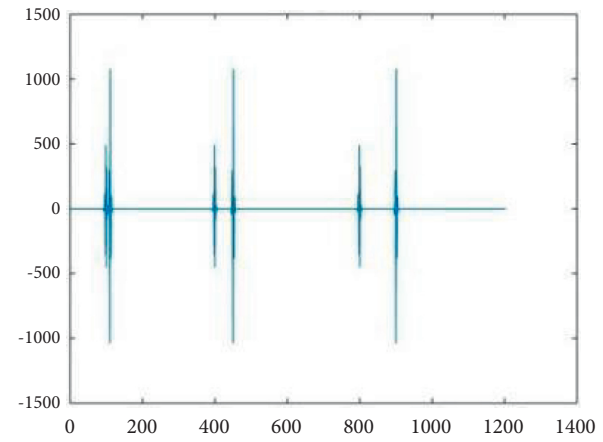

Figure 10: Detection graph.

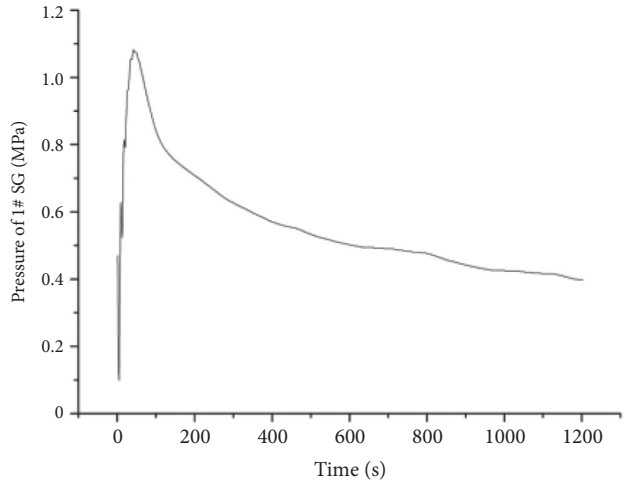

Figure 11: Curve after filling.

Table 4: Comparison of the filling error rates for missing data in the pressure of 1# SG.

No.	HD	MI	LSTM	No.	HD	MI	LSTM	No.	HD	MI	LSTM
11	3.27E-06	0.00001470	5.13E-05	51	2.29E-05	0.50000000	5.88E-05	131	1.67E-04	0.49991504	1.71E-05
12	3.95E-04	0.50000000	5.08E-05	52	2.29E-05	0.50000000	5.98E-05	132	1.67E-04	0.49991504	2.70E-05
13	6.53E-06	0.49978437	4.99E-05	53	2.29E-05	0.50000000	4.85E-05	133	1.67E-04	0.49991504	7.06E-05
14	3.27E-06	0.49978437	6.1E-05	54	2.29E-05	0.50000000	4.85E-05	134	1.67E-04	0.49991504	7.20E-05
15	6.53E-06	0.49978274	6.1E-05	55	2.29E-05	0.50000000	4.85E-05	135	1.70E-04	0.49991341	6.17E-05
16	6.53E-06	0.49978274	6.1E-05	56	2.29E-05	0.50000000	4.85E-05	136	1.70E-04	0.49991341	5.04E-05
17	3.27E-06	0.49999837	9.44E-05	57	2.61E-05	0.49999837	8.19E-05	137	1.70E-04	0.49991341	5.04E-05
18	3.27E-06	0.49999837	5.13E-05	58	2.61E-05	0.49999837	3.87E-05	138	1.70E-04	0.49991341	1.71E-05
22	6.53E-06	0.49999673	2.77E-05	62	0	0.00000163	2.74E-05	142	1.73E-04	0.49991177	6.17E-05
23	6.53E-06	0.49999673	3.75E-05	63	1.27E-04	0.50000000	3.83E-05	143	1.73E-04	0.49991177	5.05E-05
24	6.53E-06	0.49999673	4.77E-05	64	1.27E-04	0.49993465	3.73E-05	144	1.73E-04	0.49991177	5.05E-05
25	9.80E-06	0.49999510	2.55E-05	65	1.27E-04	0.49993465	8.19E-05	145	1.73E-04	0.49991177	5.05E-05
26	9.80E-06	0.49999837	3.53E-05	66	1.27E-04	0.49993465	3.87E-05	146	1.73E-04	0.49991177	5.05E-05

Table 5: (Attached table) Comparison of the filling error rates for missing data in the pressure of 1# SG.

No.	HD	MI	LSTM	No.	HD	MI	LSTM	No.	HD	MI	LSTM	No.	HD	MI	LSTM
1	6.5312E-06	0.49999673	0.00166666	41	0	0.50000000	0.00166667	81	0.00014051	0.49992975	0.001666432	121	0.00016665	0.49991667	0.001666389
2	9.7969E-06	0.49999510	0.00166665	42	3.2671E-06	0.49999837	0.00166666	82	0.00014051	0.49992975	0.001666432	122	0.00016665	0.49991667	0.001666389
3	1.6328E-05	0.49999184	0.00166664	43	3.2671E-06	0.49999837	0.00166666	83	0.00014051	0.49992975	0.001666432	123	0.00016665	0.49991667	0.001666389
4	1.3063E-05	0.49999020	0.00166663	44	3.2671E-06	0.49999837	0.00166666	84	0.00014051	0.49992975	0.001666432	124	0.00016665	0.49991667	0.001666389
5	6.5314E-06	0.49999020	0.00166663	45	3.2671E-06	0.49999837	0.00166666	85	0.00014377	0.49992811	0.001666427	125	0.00016665	0.49991667	0.001666389
6	9.7971E-06	0.49999510	0.00166665	46	3.2671E-06	0.49999837	0.00166666	86	0.00014377	0.49992811	0.001666427	126	0.00016665	0.49991667	0.001666389
7	1.6329E-05	0.49999184	0.00166664	47	3.2671E-06	0.49999837	0.00166666	87	0.00014377	0.49992811	0.001666427	127	0.00016665	0.49991667	0.001666389
8	1.9594E-05	0.49999020	0.00166663	48	6.5343E-06	0.49999673	0.00166666	88	0.00014377	0.49992811	0.001666427	128	0.00016992	0.49991504	0.001666383
9	2.6126E-05	0.49998694	0.00166662	49	6.5343E-06	0.49999673	0.00166666	89	0.00014377	0.49992811	0.001666427	129	0.00016992	0.49991504	0.001666383
10	2.9392E-05	0.49998530	0.00166662	50	0	0.49999673	0.00166666	90	0.00014704	0.49992648	0.001666422	130	0.00016992	0.49991504	0.001666383
11	3.2658E-05	0.00001470	4.9E-08	51	0	0.50000000	0.00166667	91	0.00014704	0.49992648	0.001666422	131	0.00016992	0.49991504	0.001666383
12	0.00043125	0.50000000	0.00166667	52	0	0.50000000	0.00166667	92	0.00014704	0.49992648	0.001666422	132	0.00016992	0.49991504	0.001666383
13	0.00043125	0.49978437	0.00166595	53	0	0.50000000	0.00166667	93	0.00014704	0.49992648	0.001666422	133	0.00016992	0.49991504	0.001666383
14	0.00043125	0.49978437	0.00166595	54	0	0.50000000	0.00166667	94	0.00015031	0.49992484	0.001666416	134	0.00016992	0.49991504	0.001666383
15	0.00043452	0.49978274	0.00166594	55	0	0.50000000	0.00166667	95	0.00015031	0.49992484	0.001666416	135	0.00017319	0.49991341	0.001666378
16	3.2671E-06	0.49978274	0.00166594	56	0	0.50000000	0.00166667	96	0.00015031	0.49992484	0.001666416	136	0.00017319	0.49991341	0.001666378
17	3.2671E-06	0.49999837	0.00166666	57	3.2672E-06	0.49999837	0.00166666	97	0.00015031	0.49992484	0.001666416	137	0.00017319	0.49991341	0.001666378
18	3.2671E-06	0.49999837	0.00166666	58	3.2672E-06	0.49999837	0.00166666	98	0.00015031	0.49992484	0.001666416	138	0.00017319	0.49991341	0.001666378
19	3.2671E-06	0.49999837	0.00166666	59	3.2672E-06	0.49999837	0.00166666	99	0.00015031	0.49992484	0.001666416	139	0.00017319	0.49991341	0.001666378
20	6.5342E-06	0.49999673	0.00166666	60	3.2672E-06	0.49999837	0.00166666	100	0.00015358	0.49992321	0.001666411	140	0.00017319	0.49991341	0.001666378
21	6.5342E-06	0.49999673	0.00166666	61	3.2672E-06	0.49999837	0.00166666	101	0.00015358	0.49992321	0.001666411	141	0.00017646	0.49991177	0.001666373
22	6.5342E-06	0.49999673	0.00166666	62	3.2672E-06	0.00000163	5.4333E-09	102	0.00015358	0.49992321	0.001666411	142	0.00017646	0.49991177	0.001666373
23	6.5342E-06	0.49999673	0.00166666	63	0.0001307	0.50000000	0.00166667	103	0.00015358	0.49992321	0.001666411	143	0.00017646	0.49991177	0.001666373
24	6.5342E-06	0.49999673	0.00166666	64	0.0001307	0.49993465	0.00166645	104	0.00015685	0.49992158	0.001666405	144	0.00017646	0.49991177	0.001666373
25	3.2671E-06	0.49999510	0.00166665	65	0.0001307	0.49993465	0.00166645	105	0.00015685	0.49992158	0.001666405	145	0.00017646	0.49991177	0.001666373
26	3.2671E-06	0.49999837	0.00166666	66	0.0001307	0.49993465	0.00166645	106	0.00015685	0.49992158	0.001666405	146	0.00017646	0.49991177	0.001666373
27	3.2671E-06	0.49999837	0.00166666	67	0.0001307	0.49993465	0.00166645	107	0.00015685	0.49992158	0.001666405	147	0.00017973	0.49991014	0.001666367
28	3.2671E-06	0.49999837	0.00166666	68	0.0001307	0.49993465	0.00166645	108	0.00015685	0.49992158	0.001666405	148	0.00017973	0.49991014	0.001666367
29	3.2671E-06	0.49999837	0.00166666	69	0.00013397	0.49993302	0.00166644	109	0.00015685	0.49992158	0.001666405	149	0.00017973	0.49991014	0.001666367
30	6.5342E-06	0.49999673	0.00166666	70	0.00013397	0.49993302	0.00166644	110	0.00016012	0.49991994	0.0016664	150	0.00017973	0.49991014	0.001666367
31	6.5342E-06	0.49999673	0.00166666	71	0.00013397	0.49993302	0.00166644	111	0.00016012	0.49991994	0.0016664	151	0.00017973	0.49991014	0.001666367
32	6.5342E-06	0.49999673	0.00166666	72	0.00013397	0.49993302	0.00166644	112	0.00016012	0.49991994	0.0016664	152	0.00018299	0.49990850	0.001666362
33	6.5342E-06	0.49999673	0.00166666	73	0.00013397	0.49993302	0.00166644	113	0.00016012	0.49991994	0.0016664	153	0.00018299	0.49990850	0.001666362
34	6.5342E-06	0.49999673	0.00166666	74	0.00013397	0.49993302	0.00166644	114	0.00016338	0.49991831	0.001666394	154	0.00018299	0.49990850	0.001666362
35	6.5342E-06	0.49999673	0.00166666	75	0.00013724	0.49993138	0.00166644	115	0.00016338	0.49991831	0.001666394	155	0.00018299	0.49990850	0.001666362
36	9.8014E-06	0.49999510	0.00166665	76	0.00013724	0.49993138	0.00166644	116	0.00016338	0.49991831	0.001666394	156	0.00018299	0.49990850	0.001666362
37	9.8014E-06	0.49999510	0.00166665	77	0.00013724	0.49993138	0.00166644	117	0.00016338	0.49991831	0.001666394	157	0	0.49990850	0.001666362
38	9.8014E-06	0.49999510	0.00166665	78	0.00013724	0.49993138	0.00166644	118	0.00016338	0.49991831	0.001666394	158	0	0.50000000	0.001666667
39	0	0.49999510	0.00166665	79	0.00013724	0.49993138	0.00166644	119	0.00016338	0.49991831	0.001666394	159	3.2677E-06	0.49999837	0.001666661
40	0	0.50000000	0.00166667	80	0.00014051	0.49992975	0.00166643	120	0.00016338	0.49991831	0.001666394	160	3.2677E-06	0.49999837	0.001666661

Table 6: Comparison of the average error rates.

	Parameter	HD	Mean interpolation	LSTM
Discrete missing	Outlet temperature of left loop	3.63E-06	6.16E-06	4.48E-05
	Pressure of PRZ	1.10E-05	5.00E-02	2.20E-04
Continuous missing	Pressure of 1# SG	4.86E-05	0.491891	0.00164

Table 7: The calculating time of the designed program.

No.	Time (s)	No.	Time (s)	No.	Time (s)	No.	Time (s)	No.	Time (s)
1	0.061	11	0.047	21	0.015	31	0.062	41	0.044
2	0.031	12	0.046	22	0.016	32	0.071	42	0.063
3	0.047	13	0.042	23	0.046	33	0.048	43	0.07
4	0.032	14	0.05	24	0.047	34	0.03	44	0.048
5	0.068	15	0.071	25	0.053	35	0.046	45	0.03
6	0.065	16	0.032	26	0.031	36	0.03	46	0.014
7	0.016	17	0.03	27	0.016	37	0.048	47	0.042
8	0.063	18	0.047	28	0.045	38	0.03	48	0.031
9	0.062	19	0.045	29	0.047	39	0.031	49	0.063
10	0.016	20	0.016	30	0.03	40	0.046	50	0.043

The missing data is filled with the designed algorithm, and the results are shown in Figure 11. A part of the error rates of the calculation results are listed and compared with the other two algorithms in Table 4. And the full vision of the data is listed in the Attached Table 5.

7. Result Analysis and Conclusion

The average error rates are listed in Table 6 to evaluate the stability of the algorithms from another perspective.

By analyzing the results of the experiment result in Section 5, the following conclusion could be drawn.

(1) In the experiment on the data with discrete missing points, the error rates of the HD algorithm are commonly lower than those of the other methods.

(2) In the experiment on the pressure of PRZ, the average error rate of the MI method rises sharply, while the other two methods raised slightly. By analyzing the original data, it can be found that some of the discrete missing points are located continuously. Then it can be further concluded that the HD algorithm is much more stable. Moreover, in the experiment on the data with continuous missing, this conclusion is proved more clearly.

(3) Generally, the designed HD algorithm based on Mahalanobis distance performs better than other origin algorithm both on accuracy and on stability.

In the process of experimental verification, 50 data points were randomly selected to record their calculating time through the program, and the obtained list is shown in Table 7.

The average of the time in the table is 0.04246 seconds. It indicates that the designed algorithm is of very high efficiency (considering that the experiment has been carried out on a normal PC). And since the average calculating time is far less than the data acquisition frequency, the designed algorithm is also fit for online applications.

In summary, the missing data filling algorithm of NPP is studied based on analyzing the characteristics of NPP operation data. The missing data detection method based on wavelet decomposition is studied to identify the normal zero value and data missing, which solves the problem of unclear criteria for data missing. The construction method for operation state vector of NPP is studied. On this basis, the similarity measurement of the analog parameter vector based on Mahalanobis distance, the similarity measurement of the switch parameter vector based on match measure, as well as their joint similarity measure are studied. Then the entire algorithm flow of missing data filling algorithm for NPP is designed. Finally, the designed algorithm is verified by experiments, which proves its correctness and feasibility. And it performs better than some commonly used algorithms.

The application prospect of the designed algorithm may lay on the following aspects.

(1) For the NPP operation data offline analysis, it can serve for data cleaning before the application of big data analysis for NPP abnormal operation state detection and operation experience feedback, to improve data quality and optimize data analysis results.

(2) For the NPP operation data online analysis, it can be used to correct the measurement error when the sensor or the measuring channel fails, to improve the function of fault tolerant control.

Data Availability

The data used to support the findings of this study were supplied by Chen Yusheng under license and so cannot be made freely available. Requests for access to these data should be made to [Chen Yusheng, 1402590869@qq.com].

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] M. Chmielewski and S. C. Kucker, "An MTurk crisis? Shifts in data quality and the impact on study results," *Social Psychological and Personality Science*, vol. 11, no. 4, pp. 464–473, 2020.

[2] N. Zhang and Q. Yuan, "A review of data quality evaluation," *Information Studies: Theory & Application*, vol. 40, no. 10, pp. 135–139, 2017.

[3] T. Rekatsinas, X. Chu, I. F. Ilyas, and C. Ré, "HoloClean," *Proceedings of the VLDB Endowment*, vol. 10, no. 11, pp. 1190–1201, 2017.

[4] I. F. Ilyas and X. Chu, "Trends in cleaning relational data: consistency and deduplication," *Foundations and Trends in Databases*, vol. 5, no. 4, pp. 281–393, 2015.

[5] L. O. Silva and L. E. Zárate, "A brief review of the main approaches for treatment of missing data," *Intelligent Data Analysis*, vol. 18, no. 6, pp. 1177–1198, 2014.

[6] I. Fortes, L. Mora-López, and F. Triguero, "Inductive learning models with missing values," *Mathematical and Computer Modelling*, vol. 44, no. 9-10, pp. 790–806, 2006.

[7] J. R. Quinlan, *Programs for Machine Learning*, Morgan Kaufmann Series in Machine Learning, São Francisco, CA, 1993.

[8] R. J. A. Little and D. B. Rubin, *Statistical Analysis with Missing Data*, Wiley, NJ, USA, Series in Probability and Statistics, 2nd Edition, 2002.

[9] A. Ragel and B. Crémilleux, "Treatment of missing values for association rules," in *Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pp. 258–270, Melbourne, Australia, 1998.

[10] A. Ragel and B. Crémilleux, "MVC-A preprocessing method to deal with missing balues," *Knowledge-Based Systems*, vol. 12, no. 5–6, pp. 285–291, 1999.

[11] A. Revenko, O. S. Kuznetsovm, and B. Ganter, *Finding Errors in New Object Intents, Proc CLA*, L. Szathmary and U. Priss, Eds., pp. 151–162, Universidad de Málaga Spain, Málaga, Spain, 2012.

[12] J. A. Soares, *Pré-processamento em mineração de dados: Um estudo comparativo em complementação, Doctoral dissertation*, Federal University of Rio de Janeiro, COPPE, Rio de Janeiro, RJ, Brasil, 2007.

[13] J. Luengo, S. García, and F. Herrera, "A Study on the use of imputation methods for experimentation with radial basis function network classifiers handling missing attribute values: the good synergy between RBFs and Event Covering method," *Neural Networks*, vol. 2, no. 3, pp. 406–418, 2010.

[14] C. Wu, C. Wun, and H. Chou, "Using association rules for completing missing data," in *Proceedings of the Fourth IEEE International Conference on Hybrid Intelligent System, 2004 (HIS'04)*, pp. 236–241, Kitakyushu, Japan, December 2004.

[15] H. Cui, "A novel advancing signal processing method based on coupled multi-stable stochastic resonance for fault detection," *Applied Sciences*, vol. 11, pp. 5385–5398, 2021.

[16] S. Gao, Y. G. Tang, and X. Qu, "LSSVM Based Missing Data Imputation in Nuclear Power Plant's Environmental Radiation Monitor Sensor Network," in *Proceedings of the 2012 IEEE fifth International Conference on Advanced Computational Intelligence (ICACI)*, pp. 479–484, Nanjing, China, 2012.

[17] X. Ran, "A novel K-means clustering algorithm with a noise algorithm for capturing urban hotspots," *Applied Science*, vol. 11, no. 23, Article ID 11202, 2021.

[18] Z.-H. Zhang, "Tri-partition state alphabet-based sequential pattern for multivariate time series," *Applied Sciences*, vol. 11, pp. 11202–11221, 2021.

[19] O. Azeroual, G. Saake, and J. Wastl, "Data measurement in research information systems: metrics for the evaluation of data quality: metrics for the evaluation of data quality," *Scientometrics*, vol. 115, no. 3, pp. 1271–1290, 2018.

[20] R. J. A. Little and D. B. Rubin, *Statistical Analysis with Missing Data*, Wileyand Sons Inc, New York, NY, USA, 1987.

[21] S. Oba, M. A Sato, I. takemasa, M Monden, K Matsubara, and S Ishii, "A bayesian missing value estimation method for gene expression profile data," *Bioinformatics*, vol. 19, no. 16, pp. 2088–96, 2003.

[22] G. E. A. P. A. Batista and M. C. Monard, "An analysis of four missing data treatment methods for supervised learning," *Applied Artificial Intelligence*, vol. 17, no. 5-6, pp. 519–533, 2003.

[23] M. Magnani, "Techniques for dealing with missing data in knowledge discovery tasks," *Department of Computer Science*, 2003.

[24] Z. Zhang, L. Yu, and J. Deng, "Application and analysis of wavelet in detecting discontinuous points," *Electronic Technology*, vol. 39, no. 11, pp. 50–52, 2012.